

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2001, 23:43:20 ; Search time 3991.74 Seconds
(without alignments)
2661.333 Million cell updates/sec

Title: US-09-602-597-3
Perfect score: 1516
Sequence: 1 cgtcgcgagcctgggggaaa.....aaaaaaaaaaaaaaaaaa 1516

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp1:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_esta1:*
88: gb_esta2:*
89: gb_esta3:*
90: gb_esta4:*
91: gb_esta5:*
92: gb_esta6:*
93: gb_esta7:*
94: gb_esta8:*
95: gb_esta9:*
96: gb_esta10:*
97: gb_esta11:*
98: gb_esta12:*
99: gb_esta13:*
100: gb_esta14:*
101: gb_esta15:*
102: gb_esta16:*
103: gb_esta17:*
104: gb_esta18:*
105: gb_esta19:*
106: gb_esta20:*
107: gb_esta21:*
108: gb_esta22:*
109: gb_esta23:*
110: gb_esta24:*
111: gb_esta25:*
112: em_estum21:*
113: em_estum22:*
114: em_estum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16: *
118: em_estp17: *
119: em_estp18: *
120: em_estp19: *
121: em_estp20: *
122: em_estp21: *
123: em_estp22: *
124: em_estp23: *
125: em_estp24: *
126: em_estp25: *
127: em_estp26: *
128: em_estp27: *
129: em_estp28: *
130: em_estp29: *
131: em_estp30: *
132: em_estp31: *
133: em_estp32: *
134: em_estp33: *
135: em_estp34: *
136: em_estp35: *
137: em_estp36: *
138: em_estp37: *
139: em_estp38: *
140: em_estp39: *
141: em_estp40: *
142: em_estp41: *
143: em_estp42: *
144: em_estp43: *
145: em_estp44: *
146: em_estp45: *
147: em_estp46: *
148: em_estp47: *
149: em_estp48: *
150: em_estp49: *
151: em_estp50: *
152: em_estp51: *
153: em_estp52: *
154: em_estp53: *
155: em_estp54: *
156: em_estp55: *
157: em_estp56: *
158: em_estp57: *
159: em_estp58: *
160: em_estp59: *
161: em_estp60: *
162: em_estp61: *
163: em_estp62: *
164: em_estp63: *
165: em_estp64: *
166: em_estp65: *
167: em_estp66: *
168: em_estp67: *
169: em_estp68: *
170: em_estp69: *
171: em_estp70: *
172: em_estp71: *
173: em_estp72: *
174: em_estp73: *
175: em_estp74: *
176: em_estp75: *
177: em_estp76: *
178: em_estp77: *
179: em_estp78: *
180: em_estp79: *
181: em_estp80: *
182: em_estp81: *
183: em_estp82: *
184: em_estp83: *
185: em_estp84: *
186: em_estp85: *
187: em_estp86: *
188: em_estp87: *
189: em_estp88: *

190: gb_gss25: *
191: gb_gss26: *
192: gb_gss27: *
193: gb_gss28: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|-----|----------|---------------------|
| 1 | 642.8 | 42.4 | 954 | 109 | BE563252 | BE563252 601335555 |
| 2 | 602.8 | 39.8 | 699 | 109 | BE514802 | BE514802 6013355734 |
| 3 | 587.8 | 38.8 | 839 | 135 | BE742339 | BE742339 601335575 |
| 4 | 541.8 | 35.7 | 718 | 38 | AV708191 | AV708191 AV708191 |
| 5 | 534.8 | 35.3 | 539 | 25 | AI801867 | AI801867 tx28ec3.x |
| 6 | 492.4 | 32.5 | 494 | 111 | BE672806 | BE672806 7d24d08.x |
| 7 | 483 | 31.9 | 495 | 26 | AI870611 | AI870611 v175ed2.x |
| 8 | 479 | 31.6 | 479 | 27 | AI991347 | AI991347 wu42h12.x |
| 9 | 473 | 31.2 | 489 | 14 | AA995943 | AA995943 0098a07.s |
| 10 | 467.4 | 30.8 | 469 | 91 | AW471404 | AW471404 xw62901.x |
| 11 | 465 | 30.7 | 466 | 23 | AI683914 | AI683914 tw54f12.x |
| 12 | 459 | 30.3 | 459 | 26 | AI928415 | AI928415 w67e11.x |
| 13 | 450.6 | 29.7 | 457 | 13 | AA909585 | AA909585 0118f12.s |
| 14 | 440.4 | 29.1 | 442 | 23 | AI683813 | AI683813 tx77d10.x |
| 15 | 424.4 | 28.0 | 426 | 90 | AW468787 | AW468787 hc28a08.x |
| 16 | 418.8 | 27.6 | 457 | 13 | AA918424 | AA918424 0170h01.s |
| 17 | 406 | 26.8 | 406 | 19 | AI359905 | AI359905 qy35b03.x |
| 18 | 404 | 26.6 | 412 | 90 | AW469186 | AW469186 hc78q08.x |
| 19 | 401.4 | 26.5 | 411 | 19 | AI358220 | AI358220 gw14a04.x |
| 20 | 393 | 25.9 | 420 | 15 | AI050920 | AI050920 0w32h11.s |
| 21 | 384 | 25.3 | 441 | 141 | H30310 | H30310 y077h04.t1 |
| 22 | 374.6 | 24.7 | 387 | 15 | AI028514 | AI028514 0w44b04.x |
| 23 | 360.2 | 23.8 | 365 | 136 | BE857931 | BE857931 7f72b08.x |
| 24 | 354.2 | 23.4 | 388 | 143 | RI6135 | RI6135 ya51h07.s2 |
| 25 | 350.4 | 23.1 | 551 | 23 | AI688299 | AI688299 w687h10.x |
| 26 | 348 | 23.0 | 387 | 143 | RI6134 | RI6134 ya51h06.s2 |
| 27 | 341 | 22.5 | 374 | 24 | AI739570 | AI739570 w135b07.x |
| 28 | 328 | 21.6 | 329 | 27 | AI991652 | AI991652 w15c08.x |
| 29 | 320.6 | 21.1 | 371 | 38 | AV708918 | AV708918 AV708918 |
| 30 | 319.8 | 21.1 | 323 | 91 | AW516060 | AW516060 x160a01.x |
| 31 | 318.4 | 21.0 | 328 | 5 | AA302987 | AA302987 EST113187 |
| 32 | 310 | 20.4 | 310 | 13 | AA918497 | AA918497 0173h44.s |
| 33 | 305.8 | 20.2 | 461 | 143 | RI6034 | RI6034 ya51h08.t2 |
| 34 | 294.6 | 19.4 | 311 | 140 | D25700 | D25700 HUNG504066 |
| 35 | 272 | 17.9 | 272 | 147 | Z38289 | Z38289 HSC05E042.n |
| 36 | 270.8 | 17.9 | 281 | 147 | Z42004 | Z42004 HSC05E041.n |
| 37 | 266.8 | 17.6 | 271 | 5 | AA298803 | AA298803 EST114404 |
| 38 | 266 | 17.5 | 306 | 111 | BE702436 | BE702436 RC5-1N106 |
| 39 | 261.2 | 17.2 | 322 | 111 | BE702431 | BE702431 RC5-1N106 |
| 40 | 260.8 | 17.2 | 317 | 91 | AW480865 | AW480865 33897.MAR |
| 41 | 255.2 | 16.8 | 475 | 143 | RI6035 | RI6035 ya51h07.t2 |
| 42 | 229.2 | 15.1 | 939 | 135 | BE735910 | BE735910 601335285 |
| 43 | 215.4 | 14.2 | 250 | 5 | AA302929 | AA302929 EST113185 |
| 44 | 187.4 | 12.4 | 245 | 89 | AW345513 | AW345513 15033.MAR |
| 45 | 170.4 | 11.2 | 172 | 140 | F01529 | F01529 HSC05G152.n |

ALIGNMENTS

RESULT 1
BE563252 954 bp mRNA
LOCUS 601335555F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3689653 5',
DEFINITION BE563252 mRNA sequence.
ACCESSION BE563252
VERSION BE563252.1 GI:9807064
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 954)
AUTHORS
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: InCyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LCM384 row: m column: 14
High quality sequence start: 18
High quality sequence stop: 737.
Location/Qualifiers

FEATURES

source

1..954
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3689653"
/clone.lib="NIH_MGC_39"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."
BASE COUNT 202 a 284 c 266 g 202 t
ORIGIN

Query Match 42.4%; Score 642.8; DB 109; Length 954;
Best Local Similarity 94.7%; Pred. No. 4.3e-124;
Matches 744; Conservative 0; Mismatches 27; Indels 15; Gaps 7;
QY 411 taatgattttaacatcagcgccacggttcttaacatcagcccttcacgtctctg 470
DB 14 TAATGATCTTTAATACATCAGCGCCACCGCTTCTTACATCACCGCTTCATCGCTGCTG 73
QY 471 cggcaattgacatgacatccctgaaggagcgccgaccttaacacagcgcggtgct 530
DB 74 CGGCACTTGACCTGACATCCTTGAAGGGGACCGGCTTATACAGCGCGGCTGCT 133
QY 531 cgtcttgctgctgttgatgatcgcgtatgagtgagtgctcttcacgtaccag 590
DB 134 CGTTCTT--GCGTGTGATGATGATGCGCTATGAGTGAAGTGCCTTCTTACAGTACAG 191
QY 591 cctggagagagagtagcagcaaatgctggcgccacagctcagatgctgcgctatg 650
DB 192 CCGGGAGAGAGTAGGACAGCAATGCGGCCACAGTCAAGATGCTGGCGGTATGCTTAA 251
QY 651 ccaacctgtccacagcgccctctgaggctgaagcgccgctgggtacagagagaggt 710
DB 252 CCACCTGTGCGACGGCCCTCTGGGGCTGAAAGCCCGCTGGGTACAGAGGAGGTCA 311
QY 711 ccctgaagcctgaagcctgaggagccttcgtgagtgacgccaacagagagctcat 770
DB 312 CCCTGAAGAGCTGAAGCTGGGGAGCCCTGGTGAAGTCAAGCCCAAGGAGGACATTTG 371
QY 771 ctcctctgcgcgtcagacataagctctcaacagcgctagaagcagcgagctgag 830
DB 372 CTCCTCTTGCCTGCGCTAGACATTAAGCTTCACAGCCCTTAAGGAGAGGCGCCAGGCTGGC 431
QY 831 aggcactctgcgcttgagagagcgcaactgctgagacctcttcacatccacctattcag 890

|||||
DB 432 AGGATCTCGGCTTGAGAGAGGCCAACTGCGAGACCTCTTCATCCACCCCTATTTCAG 491

QY 891 tgaagatgacggggagctgaggtctgtctgtcctgtcctttagagagcttcagcgt 950

DB 492 TGGAAAGATGACGGGGGATCTGAGAGCTGTGCTTGCCTTGTCTTAAAGAGACTTCACGCT 551

QY 951 ccaagactggggccaccctctcaccagcactaaatgacactaacagagctccagcct 1010

DB 552 CCAGAGCT--GGGCCACCTCTTCACGACACTAAATGCACTAAACAGGACTCCAGA--CT 609

QY 1011 gcaagccacagaccgcgctgataaagcctaac--aagcaacagctagacccttctt 1069

DB 610 GCAGCCCCAAGAC--GGCTGATTAATACCTTAACAAACAAACACACTTACACTTATGCTTT 667

QY 1070 gtccagagagctgagcaagctggtgaaacacctctcctccttaaacacgcttcaa 1129

DB 668 GTTCCAGAGAGAGCTGAGCAAGCTGTGTAACCACTCTCTCTTAAACACGTTT---- 723

QY 1130 ccaacctctcctgagcgaacctgtaaaaagtggttgatctgacagatgtcttc 1189

DB 724 -CAACAACTCTCTGAGGCCAACTGTAAACA---GTGGTGTGCTGACACCACTGTCTCC 779

QY 1190 cctccc 1195
DB 780 CTCCC 785

RESULT 2
BE514802 699 bp mRNA EST 07-AUG-2000
LOCUS 601316734F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3653345 5'

DEFINITION mRNA sequence.
ACCESSION BE514802
VERSION BE514802.1 GI:9722016
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 699)
AUTHORS
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: InCyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM326 row: c column: 10
High quality sequence stop: 660.
Location/Qualifiers

FEATURES

source

1..699
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3633345"
/clone.lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 106 a 221 c 224 g 148 t
ORIGIN

| QY | 1116 | aaacacggtttcaaccacccctctcccttggagccaacctgtataaagvtggttattgctg | 1175 |
|------------|--|---|-------------|
| Db | 181 | AAACACCGCTTTCAACCAACCTCTCCCTGGAGCCAAACCTGTAAAAAGTGGTTGATGCTG | 240 |
| QY | 1176 | acagatagttctccctccctccgcatltaagacataccggttaccgaagcaaacag | 1235 |
| Db | 241 | ACAGATAGTGTCTCCCTCCCTCCGCAATTTAGACATACAGTACTGAAAGCAAAATCAGTT | 300 |
| QY | 1236 | taagtattctcagtgctgtaaaagccgctccagtgllcttcccttccccaagcttc | 1295 |
| Db | 301 | TAAAGATTTTCATCATGCTGAAAGCCGTCCAGGTTTCCTCCCTTCCCAAGCTTC | 360 |
| QY | 1296 | tctgtaatactcccttgggaggaagctaacaatcggtgcttcccccagacctgtactagg | 1355 |
| Db | 361 | TCTGTAAATACCTCCCTTTGGCGAAGCTCAACATCGGTGCTTCCGACCTTCTGACTAGG | 420 |
| QY | 1356 | caactaggacgcaaaagaggaggaagcaagacctgttccctggagagttgtcatgtgtt | 1415 |
| Db | 421 | CACATGGGACCCAAAGGAGGAGAACCAAGGCCCTTGCGCAGTGTGATGTGCTT | 480 |
| QY | 1416 | ggtgtgacctgttatttttttaataaataaagaatgagagaattaaaaaaa | 1475 |
| Db | 481 | GGTGTGACTGTGTTATTTTATTAATAATAATAAGATGAGAGAAATTAATTGGAAAAA | 540 |
| QY | 1476 | aaaaaaaaaaaaa | 1489 |
| Db | 541 | AAAAAAAAAAAAA | 554 |
| RESULT | 5 | | |
| LOCUS | AI801867 | | |
| DEFINITION | tx8be03.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2270908 3', | EST | 16-DEC-1999 |
| KEYWORDS | mRNA sequence. | | |
| ACCESSION | AI801867 | | |
| VERSION | AI801867.1 | GI:5367339 | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo. | | |
| AUTHORS | 1 (bases 1 to 539) | | |
| TITLE | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. | | |
| JOURNAL | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), | | |
| COMMENT | Tumor Gene Index | | |
| | Unpublished (1997) | | |
| | Contact: Robert Strausberg, Ph.D. | | |
| | Tel: (301) 496-1550 | | |
| | Email: Robert.Strausberg@nih.gov | | |
| | Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. | | |
| | Emmert-Buck, M.D., Ph.D. | | |
| | CDNA Library Preparation: M. Bento Soares, Ph.D. | | |
| | CDNA Library Arrayed by: Greg Lennon, Ph.D. | | |
| | DNA Sequencing by: Washington University Genome Sequencing Center | | |
| | Clone distribution: NCI-CGAP clone distribution information can be | | |
| | found through the I.M.A.G.E. Consortium/LLNL at: | | |
| | www-bio.llnl.gov/bdip/image/image.html | | |
| | Insert Length: 673 Std Error: 0.00 | | |
| | Seq primer: -400P from Gibco | | |
| | High quality sequence stop: 470. | | |
| FEATURES | Location/Qualifiers | | |
| SOURCE | 1..539 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /clone="IMAGE:2270908" | | |
| | /clone_lib="NCI_CGAP_Lu24" | | |
| | /tissue_type="carcinoid" | | |
| | /lab_host="DH10B" | | |
| | /note="Organ: Lung; Vector: pUT73D-Pac (Pharmacia) with a | | |
| | modified polylinker; Plasmid DNA from the normalized | | |
| | library NCI_CGAP_Lu5 was prepared, and ss circles were | | |
| | made in vitro. Following HAP purification, this DNA was | | |

used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clonoids
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo.

BASE COUNT 127 a 158 c 124 g 129 t 1 others
ORIGIN

Query Match 35.3%; Score 534.8; DB 25; Length 539;
Best Local Similarity 99.4%; Pred. No. 1.4e-101;
Matches 536; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 875 catccccccttaagtggaagatgacggggagatctgagctgtgtctctgtctt 934
DB 1 CATCCCCCTTATTCAGTGAAGATGACGGGGATGTGAGCTGTCTCTGCTTGTCTT 60
QY 935 tagaggaactcagctcgaagactgaggccaccccttcaccagactaaatgactaa 994
DB 61 TAGAGGACTTCAGCGCTCCAGACTGGGGCCACCTTCTCACAGCACTAAATGACTAA 120
QY 995 caaaggaactcgaagctgagcccgccgagctgagctgagctgagctgagctgagct 1054
DB 121 CAAGGACTTCAGACTGACGCCCCAGACCCGCGATGATTAAGCTTAACAAGCAACAGT 180
QY 1055 agcaactagctgtgtgtccagagagctgagcaagctgtgtgaaacacactctcttct 1114
DB 181 AGCACTTAGTCTTGTGTCCAGAGAGCTGAGCAAGCTGTGAAACCACTCTCTCTTCT 240
QY 1115 taacaacgcttcaacaacactctccctgagagcaactgtgaaagtgggtgtgtgt 1174
DB 241 TAAACACCGTTCACCAACCACTCTCTGAGCCAACTGTAAAGAGGGGTGATTTCT 300
QY 1175 gacagcaatggtctccctccctcctgcatctcagacataccagcttctgaaagaatcagtt 1234
DB 301 GACAGCATGGCTGCTGCTCCCTGCTGATTTCAACATACCACTTCTGAAAGCAATCACTT 360
QY 1235 ttaagtgatctcagtgagtgaaagcctgtccaggttctctcccttcccaagcctct 1294
DB 361 TTAAGTGTATTCAGTCTGCTGAAAGCCTGTCCAGGTTCTCTCCCTTCCCAAGCCTCT 420
QY 1295 ctctgttaactcccttggcgagagctaacatcagtgctctcccgactgtgtgactag 1354
DB 421 CTCTGTATTAATCTCCCTTGGGCGAAGCTAATCATGCTGCTCCCGACCTTGCTACTAG 480
QY 1355 gcaatggaagcgaag 1413
DB 481 GCACATGGGACGCANAGAGGAGGAGGAGCAAGGCTTGCTGGGAGGTGTGATGTGG 539

RESULT 6 BE672806 494 bp mRNA EST 08-SEP-2000
LOCUS BE672806/c 7024a08.x1 NCI_CGAP_Pt28 Homo sapiens cDNA clone IMAGE:3248150 3'
DEFINITION similar to contains Alu repetitive element.; mRNA sequence.
ACCESSION BE672806
VERSION BE672806.1 GI:10033347
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 494)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncigap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmer-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL, send email to:
info@image.lnl.gov
Seq primer: -400p from Gibco
High quality sequence stop: 463.
Location/Qualifiers

FEATURES

source

1. 494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3248150"
/clone_id="NCI_CGAP_Pt28"
/sex="male"
/dex_stage="adult"
/lab_host="DH10B"

/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pt22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
983608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 128 a 107 c 130 g 129 t
ORIGIN

Query Match 32.5%; Score 492.4; DB 11; Length 494;
Best Local Similarity 99.8%; Pred. No. 9.3e-93;
Matches 493; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 971 tctaacgacctaataatcactaacgaagactcgaagctgaccccgagcccgctga 1030
DB 494 TCTCACGACGCTAATGCTAATGCTAATGCTAATGCTAATGCTAATGCTAATGCTAAT 435
QY 1031 gtaaaagcctaaagcaagcaagcagctagcacttctgttccagagagctgagcaag 1090
DB 434 GTATTAAGCTTAACAAGCAACAGCTAGACCTTATGCTTGTGTCCAGAGAGCTAGAGCA 375
QY 1091 ctgtgaaacacactctctctctttaaaccacgcttcaaccaactctcctgtgagcaa 1150
DB 374 CTGTGAACACGACTCTCTCTCTTAAACACGCTTCAACCAACCTCTCTCTGAGCCAA 315
QY 1151 cctgtaaaagtgtgtatgtgtgtgagagagctgtcctcctgtgattcagacta 1210
DB 314 CTGTAAAGGTGGTGTATGCTGACGACATGGCTTCTCTCTCTGATTTCAAGACTA 255
QY 1211 ccagttaagcaagcaatcagtttaagttatctcagtgctgaaagctgtccag 1270
DB 254 CCAGTTACTGAAGCAATGATGTTTAAGTATTTCTGAGTGTGAAAGCTGTGCCAG 195
QY 1271 ttctctcccttcccaagcctctctgttaatactcctcttggcgagagtaacatcg 1330
DB 194 TTTCT 135
QY 1331 tgcctcccgacactgtgtgactagacatgagcgaagagagagagagagagagag 1390
DB 134 TGCTCTCCCGACTTGTGCTGACTAGGACATGAGGAGGAGGAGGAGGAGGAGGAGGAG 75
QY 1391 ttgcctgagagctgtcagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1450
DB 74 TTGCTGTGCGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 15
QY 1451 agatgagagaaatc 1464
DB 14 AGATGAGAGAAATT 1

RESULT 7 A1870611 495 bp mRNA EST 07-MAR-2000
LOCUS A1870611/c

| Query Match | 31.6% | Score 479; | DB 27; | Length 479; |
|-------------|-------|------------|--------|-------------|
| BASE COUNT | 125 a | 105 c | 125 g | 124 t |
| ORIGIN | | | | |
| Query Match | | | | |

31.68; Score 479; DB 27; Length 479;

Best Local Similarity 100.0%; Pred. No. 5.8e-90;
Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 987 tgcactaacagagctccagacgctccagacccgagcgcgtgataaagcctaagaag 1046
Db 479 TGCACTAACAGAGCTCCAGACCTGCACCCGCGCGGTAGTATAAGCTTAACAG 420
Qy 1047 caaacagtagaaccttagctcttccagagagctagcaagctggtgaacacact 1106
Db 419 CAACAGTAGAACCTTAGCTCTTGTCCAGAGAGCTAGCAAGCTGTGAACACACT 360
Qy 1107 ccttcctttaaaccgcttcaaccaactctccctgagcacaactgtaaaagtgggt 1166
Db 359 CCTTCCTTTAAACACCGCTTACCAACCTCTCCCTGAGCAACCTGTAAAGAGTGGGT 300
Qy 1167 tgcattgcgaagcagatgctctccctccctcatttcagacataccagttactgaaga 1226
Db 299 TGATTGCTGACAGATGCTCTTCCCTCCCTCATTTCAGACATACCACTTACGAAACGA 240
Qy 1227 aatcagtttaagatcttcaagtgctgaagaagcctgaccaggttccctccctcc 1286
Db 239 AATCAGTTTAAAGATTTTCAGTGGCGAAGACCTGTCAGGTTTCCCTTCCCTTCC 180
Qy 1287 aagcctctctctgaatactcccttggggagagctaacatcggtgctcccgaccttg 1346
Db 179 AAGCTCTCTCTGTAATACCTCTTGGGCGAAGCTTAACATCGGTGCTCCCGACCTTG 120
Qy 1347 ctgactgagcagctgag 1406
Db 119 CTGACTGAGCAGATGAGGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Qy 1407 catgt 1465
Db 59 CATGT 1

RESULT 9 AA95943 489 bp mRNA EST 27-AUG-1998
LOCUS 0098807.s1 NCI_CGAP_Kid3 Homo sapiens CDNA clone IMAGE:1635828 3'
DEFINITION similar to contains Alu repetitive element; mRNA sequence.

ACCESSION AA95943
VERSION AA95943.1 GI:3182432
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 489)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

CDNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

www-bio.lnl.gov/bdip/image/image.html

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 469.

Location/Qualifiers

1. 489
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1635828"
/clone_1ib="NCI_CGAP_Kid3"

/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."

BASE COUNT 129 a 107 c 125 g 128 t
ORIGIN

Query Match 31.2%; Score 473; DB 14; Length 489;
Best Local Similarity 99.8%; Pred. No. 1e-88;
Matches 484; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 981 actaatgactaacagagctccagacctgcaagcccgagcgcgtgataaagcct 1040
Db 489 ACTAATGCACTAACAGAGCTCCAGACCTGCAGCCCGCGGTAGTATAAGGCT 430
Qy 1041 aacaagaacaagtagaaccttagctcttccagagagctgagcaagctggtgaac 1100
Db 429 AACAGCAACAGCTAGACCTTAGCTTGTTCAGAGAGACTAGCAAGCTGTGAAC 370
Qy 1101 caactcctctctttaaaccgcttcaaccaactctccctgagcacaactgtaaaaa 1160
Db 369 CACTCTCTCTCTTTAAACACCGTTTCAACACCACTCTCCCTGAGCCACCTGTAAANA 310
Qy 1161 gtgggttgatgtctgacagcagctgctccctccctcatttcagacatacagctactg 1220
Db 309 GTGGGTGATGTGCGAGCAGCTGCTTCCCTCCCTGATTTAGACATACCACTACTG 250
Qy 1221 aaagcaaatcagtttaagatcttcaagtgctgaagaagcctgtccaaggttccctcc 1280
Db 249 AAACCAATCAGTTTAAAGTATTTCTCAGTGTGAAAGCCTGTCCAGGTTTCTTCC 190
Qy 1281 ttcccaagcctctctctgtaatactcccttgggagagagtaacatcggtcctcccg 1340
Db 189 TTCCCAAGCCTCTCTCTGTATTAATCT-CCTTGGGCGAAGGTAACTAGGCTCCCG 131
Qy 1341 acctgtgactaagcagcatgagcagcaagagagagagagagagagagagagagag 1400
Db 130 ACCTGTGACTAGCAGCATGAGCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 71
Qy 1401 agttgtcatgt 1460
Db 70 AGTTGTCAATGT 11
Qy 1461 aatta 1465
Db 10 AATTA 6

RESULT 10 AA471404 469 bp mRNA EST 24-FEB-2000
LOCUS AM471404/c xw62g01.x1 NCI_CGAP_Pan1 Homo sapiens CDNA clone IMAGE:2832624 3'
DEFINITION similar to contains Alu repetitive element; mRNA sequence.

ACCESSION AA471404
VERSION AA471404.1 GI:7041510
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 469)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Life Technologies catalog #: 11348-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/UNL at:
www.bio.linn.gov/bdip/image/image.html
Seq primer: -400P from Gibco
High quality sequence stop: 424.

| | | | | |
|------------|-------|-------|-------|-------|
| BASE COUNT | 124 a | 104 c | 121 g | 120 t |
| ORIGIN | | | | |

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 30.8%; | Score 467.4; | DB 91; | Length 469; |
| Best Local Similarity | 99.8%; | Pred. No. 1.5e-87; | | |
| Matches 468; Conservative | 0; | Mismatches 1; | Indels 0; | Gaps 0; |

| | | | |
|----|------|--|------|
| OY | 397 | agagctccagacccgagagccagaccgcgcgtagatataagctaaacgaacacgtag | 1056 |
| Db | 469 | AGGACTCCAGACTGCGAGCCCGACGACCCGCTGTAGTATTAAGCTTAACAGCAACGATG | 410 |
| OY | 1057 | caacctagctcttgctccagagagactggaacactgctgtaaaccaactccctccctta | 1118 |
| Db | 409 | CACCTTAGTCTTGTGTTCCAGAGAGAGCTGAGCAAGCTGTGTGAACCACTCTCTCTTTTA | 350 |
| OY | 1117 | aaccgcgttccaaacaaactctccctcggagccaaactgtaaaaagtgggtattgtcgtga | 1176 |
| Db | 349 | AACACGCTTTCAACCAACCTCTCCCTGGAGCCAACTGTAAAAAGTGGGTATTATTCGTGA | 290 |
| OY | 1177 | cagcatgctctccctccctccgcatcttcagacacataccagttactgtaagaacacagcttt | 1236 |
| Db | 289 | CAGCATGCTCTTCCCTCCCTCCGCACTTCAGACATACACGTTACTGAAAGCAATCAGTTTT | 230 |
| OY | 1237 | aagtgatcttccagtgctgtaaaagcctgtccagtgctctccctcccttcccaagcctctct | 1296 |
| Db | 229 | AAGTGATTTTCTCAGTCTCTGAAAAAGCCTGTGCCAGGTTCTCTTCCCTTCCCAAGCCTCTCT | 170 |
| OY | 1297 | ctgtaataactccctcttgggcgaaagctcaacatcgtgtgctcccccagacctgtcgtactaagc | 1356 |
| Db | 169 | CTGTAAATACCTCCTTTTGGGCGAAGCTTAACATCGTGTCCTCCCGCACTTGTGCTGATTAAGC | 110 |
| OY | 1357 | acatgaggacgaagaagagagagaaacagagccttgcctctgacgagtgtgcatcgtgtg | 1416 |
| Db | 109 | ACATGGAGACCAAGAAGAGGAGGAGCAAGGCAAGGCTTGCTGCGAGTTGTCAATGCGTTG | 50 |
| OY | 1417 | gtggtgacgtgtttatttttttttaataaaaaataaagatgagagaataa | 1465 |
| Db | 49 | GTGTGTCACCTGTTTATTTTATTTTAAAAAATAAATTAACATGAGGAATTA | 1 |

| | |
|------------|--|
| RESULT | 11 |
| A1683914/c | |
| LOCUS | |
| DEFINITION | A1683914 466 bp mRNA EST 15-DEC-1999 |
| ACCESSION | tws4fi12.x1 NCI CGAP UCL Homo sapiens CDNA clone IMAGE:2263535 3' |
| VERSION | A1683914 similar to contains Alu repetitive element.; mRNA sequence. |
| KEYWORDS | A1683914.1 GI:4894096 |
| SOURCE | EST. |
| ORGANISM | human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi |

REFERENCE Mammalia: Eutheria: Primates; Catarrhini; Homnidae; Homo
AUTHORS 1 (bases 1 to 466)
TITLE NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project a
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .466 |

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 30.7%; | Score 465; | DB 23; | Length 466; |
| Best Local Similarity | 99.8%; | Pred. NO. 4.8e-87; | | |
| Matches 465; Conservative | 0; | Mismatches 1; | Indels 0; | Gaps 0; |

| | | | |
|----|------|--|------|
| QY | 1012 | caaccccaagaccgcccgtatgataaacctcaacagaagaacaglaagaccttagcttctgt | 1071 |
| QY | 1012 | caaccccaagaccgcccgtatgataaacctcaacagaagaacaglaagaccttagcttctgt | 1071 |
| Db | 466 | CAGCNCAGACCCGCCGTATATTAACCTTAACAGCAACGATGACCTTAGCTTTGT | 407 |
| QY | 1072 | tccaaggaagacttgagcaagcttggltgaaacaactctctctcttlaaaccggtlccaac | 1131 |
| Db | 406 | TCCAGAGAGACTGAGCAAGCTGGTGAACCACTCTCTCTTTAAACACGTTTCAMC | 347 |
| QY | 1132 | aaacctctcccttgaagccaactgttaaaaagtgtggttgaattgcttgacaagcatgtctccc | 1191 |
| Db | 346 | AACCTCTCCCTGGAGCCAACTGTATAAAGGTGGTTGATGTGACAGCATGCTTCC | 287 |
| QY | 1192 | tcacctgattcagacataccagcttactgtaaagcaaatcagtttlaagatcttcagt | 1251 |
| Db | 286 | TCCTCGATTTCAGACATACACATTAACCTGAAGCAAAATGATTTAAAGATTTCTGAGT | 227 |
| QY | 1252 | gctgaaaagcctgtccaggttctctccctttcccaagcctctctctglaatactccct | 1311 |
| Db | 226 | GCTGAAAAGCGTCTCCAGATTTCTCTCCCTTCCCAAGCCTCTCTGTGAATAACCTCTT | 167 |
| QY | 1312 | tggaggaagctaacatcagtgctgcctcccgcaacttctgactgagcaagtggagccgaag | 1371 |
| Db | 166 | TGGGCGAGCTTAACATTCGGTGCCTCCGCCAACCCTGTGCTAGGACATAGGACCCCAAG | 107 |
| QY | 1372 | gaaggaggaagcaagccttgctgctgtgcgaattgtcatatgtgttggltgactcttca | 1431 |
| Db | 106 | GAGGAGAGGAGCAAGCCCTTGCTGCTGGCAGATTGTCATGTGGTTGGTGTGACTTTTA | 47 |
| QY | 1432 | ttttttttataaaaataaagatgtgaggaatataaaaaaanaa 1477 | |
| Db | 46 | TTTTTTTTTATAATAAATAAAGATGACAGAAATTTAAAAAANAANAANAANA 1 | |

RESULT 12
AI928415/c 459 bp mRNA EST 08-MAR-2000
LOCUS w067e11.x1 NCI_CGAP_Pr22 Homo sapiens CDNA clone IMAGE:2460428 3'
DEFINITION similar to contains Alu repetitive element; mRNA sequence.
ACCESSION AI928415
VERSION AI928415.1 GI:5664379
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 459)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
Tumor Gene Index
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.livnl.gov/bdrrp/image/image.html
Insert Length: 1504 Std Error: 0.00
Seq primer: -400p from Glbco.
Location/Qualifiers
FEATURES
source
1..459
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2460428"
/clone_id="NCI_CGAP_Pr22"
/sex="male"
/tissue="normal prostate"
/lab_host="DH10B"
/note="Organ: prostate; Vector: p773D-Pac (Pharmacia)
with a modified polylinker; 1st strand CDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded CDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified p773 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 121 a 102 c 122 g 114 t
ORIGIN
Query Match 30.3%; Score 459; DB 26; Length 459;
Best Local Similarity 100.0%; Pred. NO. 8.5e-86;
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1001 ctccagacctgcaagcccgccgagctatataagcctaacaagacaagcgtagcacc 1060
|||||
DB 459 CTCGAGACCTGACAGCCCGGAGCCGCTAGTATTAAGCTTAACAAGACAGTAGACACC 400
|||||

QY 1061 ttagctcttgcacagagagctgagcaagctggtgaacaacctctccttctaaca 1120
|||||
DB 399 TTACTCTTTGTTCCAGGAGAGCTGAGCAAGCTGGTGAACACACTCTCTCTTAACA 340
|||||

QY 1121 ccgcttcaaccaacctctccctgagagcaacctgtaaaagtgggtgattgctgagcgc 1180
|||||
DB 339 CCGTTTCAACCAACCTCTCCCTGAGGCCAAGCTGAAAGTGGGTGATTGCTGACAGC 280
|||||

QY 1181 atggctctccctccctgacattcagacalacacagttactgaaacaacatcagtttaag 1240
|||||
DB 279 ATGGCTCTCCCTCCCTGATTTACAGACTTACAGACTTACGAAACAAATAGTTTAAGT 220
|||||

QY 1241 gattctcagtgctgaaagcctgccaaggttctccttcccttccccaagcctctctgt 1300
|||||

RESULT 13
AA909585/c 457 bp mRNA EST 23-JUN-1998
LOCUS o118f12.s1 Soares_NFL_T_GBC_S1 Homo sapiens CDNA clone
DEFINITION IMAGE:1523855 3', mRNA sequence.
ACCESSION AA909585
VERSION AA909585.1 GI:3048990
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 457)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
Tumor Gene Index
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LINL; contact the
IMAGE Consortium (infoimage.livnl.gov) for further information.
Insert Length: 691 Std Error: 0.00
Seq primer: -40m13 fwd: ET from Amersham
High quality sequence stop: 434.
Location/Qualifiers
FEATURES
source
1..457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1523855"
/clone_id="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCL-CGAP GCB1) were mixed, and as circles were
in vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 121 a 104 c 117 g 115 t
ORIGIN
Query Match 29.7%; Score 450.6; DB 13; Length 457;
Best Local Similarity 99.1%; Pred. NO. 4.8e-84;
Matches 453; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1011 gcaagcccaagaccgcccagctatataagcctaacaagacaagcgtagccttcttg 1070
|||||
DB 457 GCAAGCCCAAGACCAGCCCTGATTAAGCTTAACAAGCAAGCAGTACACCTTAGTCTTTG 398
|||||

QY 1071 ttcaagagagctgagcaagctggtgaacaacctctccttctaacaacaccttcaac 1130
|||||

|||||
Db 397 TTCAGAGAGAGCTGAGCAAGCTGTGGAACCACTCTCTTCTTAAACACCGTTTCAAC 338
Qy 1131 caacctccctggaagcaacctgtaaaagtgggtgattgttcgaaagatggtcttc 1190
Db 337 CAACCTCTCCCTGGAGCAACCTTTAAAGTGGTGTGATGTCGACAGCATGGTCTTCC 278
Qy 1191 ctccctcatttcagacatacccaattactgaagaacaatcaatttaagtatttcag 1250
Db 277 CTCCTTCATTTTCAGACATACCACTTACTGAAGCAATCAGTTTAAAGTATTTCTCAG 218
Qy 1251 tgcgtaaaagccgtgccaaggttccctccctcccaagcctctctctgtataactcct 1310
Db 217 TGCTGAAGAAGCTGTCCAGGTTTCTCTCCCTTCCCAAGCCTCTCTGTAATCTCGCT 158
Qy 1311 ttggggcaagcaatacatcggtgctcccccgaacctgtctgactaggaacatggagcaaa 1370
Db 157 TTGGGCGAAGCTTACATCACTGGTGGCTCCCGACCTTGGCTGACTAGGCAATGGAGCGCAA 98
Qy 1371 ggaaggaggaagcaagccctgctgagagttgtcatgtgtgtgtgtgtgtgttt 1430
Db 97 GGAAGGAGGAGCAAGGCGCTTGGCTTGGCAAGTTGTATGTGTGTGTGTGTGTGT 38
Qy 1431 attttttaataaaataaagatgagagaataaa 1467
Db 37 ATTTTATTAATAAATAAAGATGAGCGGCAATTAA 1
RESULT 14
A1683813/c 442 bp mRNA EST 16-DEC-1999
LOCUS tx77610.x1 NCI-CGAP-U1 Homo sapiens CDNA IMAGE:2275603 3'
DEFINITION similar to contains Alu repetitive element; mRNA sequence.
ACCESSION A1683813
VERSION A1683813.1 GI:4893995
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 442)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgaf.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Insert Length: 889 Std Error: 0.00
Seq primer: -40UP from Glibco
High quality sequence stop: 405.
Location/Qualifiers
1. 442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2275603"
/clone_lib="NCI-CGAP_U1"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

BASE COUNT 117 a 93 c 104 g 128 t

ORIGIN

Query Match 29.18; Score 440.4; DB 23; Length 442;
Best Local Similarity 99.8%; Pred. No. 6,4e-82;
Matches 441; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1050 cagctagcacttagcttctgttccagagagatcaggaacagctggtgaacacactcct 1109
Db 442 CAGCTAGCACCCTTAGCTTTTGTTCAGAGAGCTGAGCAAGCTGTGTAACCACTCTCT 383
Qy 1110 tcccttaaacacggttccaacacacccctccctggaagcaacacgttaaaagtgggtga 1169
Db 382 TCCTTTAAACACGCTTTCACACCAACCTCTCCCTGAGACCACCTGTAATAAGTGTTGA 323
Qy 1170 ttgctgaagcatggttctccctccctgcatlccagacatacagttactgaagaacat 1229
Db 322 TTGCTGACACCATGCTCTCCCTCCCTGCAATTCAGACATACAGTTATGTAAGCAAT 263
Qy 1230 cagtttaagtatttctcaagtgctgaagacgtgtccaggttctccctcccaag 1289
Db 262 CAGTTTAAGTGAATTTCTCAGTGTGTAAGCCCTGTCCAGCTTCCCTTCCCAAG 203
Qy 1290 cctctctgttaactcccttgggaggaagcaacatcggtgctcccccgaacctgtctg 1349
Db 202 CCTCTCTGTATATACCTCCCTTGGGCGAAGCTTACATCGGTGCTCCCGACCTTCTG 143
Qy 1350 actagacatggagcaaggaaggaaggaagcaagcctgtgctgaggtgtcat 1409
Db 142 ACTAGGACATGGAGGAGCAAGGAGGAGGAGCAAGGAGGAGGAGGAGGAGGAGGAGTTCAT 83
Qy 1410 gtggtgtgtgtgacgtgttattttttaataaaataaagatgagagaataaa 1469
Db 82 GTGGTGTGTGTGACCTGTATTTTAAATTAATAAATGAAGATGAGAAATTAATAAA 23
Qy 1470 aaaaaaaaaaaaaaaaaa 1491
Db 22 AAAAAAAAAAAAAAAAACAAAA 1
RESULT 15
A168787/c 426 bp mRNA EST 24-FEB-2000
LOCUS hd28a08.x1 Soares_NFL_T-GBC_S1 Homo sapiens CDNA clone
DEFINITION IMAGE:2910806 3', mRNA sequence.
ACCESSION A168787
VERSION A168787.1 GI:7038893
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 426)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgaf.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Glibco
High quality sequence stop: 377.
Location/Qualifiers
1. 426
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2910806"
/clone_lib="Soares_NFL_T-GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pVT3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;

FEATURES
source

1. 426
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2910806"
/clone_lib="Soares_NFL_T-GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pVT3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Equal amounts of plasmid DNA from three normalized libraries (fetal lung NDHL19W, testis NHT, and B-cell NCI-GAP, GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bernaldo.

BASE COUNT 117 a 94 c 104 g 111 t
ORIGIN

Query Match 28.0%; Score 424.4; DB 90; Length 426;
Best Local Similarity 99.8%; Pred. No. 1.4e-78;
Matches 425; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1043 caagcaacagctagacacttcttcttccagagagctgagcaagctggtgaacca 1102
|||||
DB 426 CAAGCAACAGCTAGCACCCTTGTTCAGAGAGCTGAGCAAGCTGGTGAACCA 367
OY 1103 ctctcccttcttaaacacgcttcaaccaactctccctgagccaaactgtaaaagt 1162
|||||
DB 366 CTCTCCTTCTTAAACACCGCTTCAACCAACCTCCATGGAGCCCAACCTGTAATAAAGT 307
OY 1163 ggggtgattgctgacagatggtctccctccctgattcagacataccaactga 1222
|||||
DB 306 GGGTGTATTGCTGACAGCATGCTCTCCCTCCCTGATTCAGACATACACTACTGAA 247
OY 1223 agcaaatcagtttaagtgattctcaagtgctgaaaagcctgtccaggttccctc 1282
|||||
DB 246 AGCAATCAGTTTAAAGTATTCAGTGTGAGAAAGCCTGTCAGGTTCTCCCTT 187
OY 1283 tcccaagcctctctgttaataactcccttggggaagtaacatcagtgccctcccgac 1342
|||||
DB 186 TCCCAAGCCTCTCTGTAAATACCTCTTGGGCGAGCTAACATCGGTGCTCCCGAC 127
OY 1343 ctgtgactagggcacatgagagcaaaagagaggaagcaagccttgctgagag 1402
|||||
DB 126 CTGTGACTAGGCACTAGGAGCGCAAGAGAGGAGCAAGCCTTGCTGGCGAG 67
OY 1403 ttgtcaatggttggtgagtgactgttattttttaataaaaataaagatgagagaa 1462
|||||
DB 66 TTGTCAATGTGTTGGTGTGACTGTTTATTTTAAATAAATAAAGATGAGAGAAA 7
OY 1463 ttaaaa 1468
|||||
DB 6 TTAATAA 1

Search completed: February 28, 2001, 23:43:31
Job time: 19397 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2001, 02:41:45 ; Search time 8425.79 Seconds
(without alignments)
920.805 Million cell updates/sec

Title: US-09-602-597-3
1516
Sequence: 1 cgtccgagcctcgtgggga.....aaaaaaaaaaaaaaaa 1516

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 118133 segs, 2558875100 residues
Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_bal:*
2: gb_bal:*
3: gb_bal:*
4: gb_bal:*
5: gb_bal:*
6: gb_bal:*
7: gb_bal:*
8: gb_bal:*
9: gb_bal:*
10: gb_bal:*
11: gb_bal:*
12: gb_bal:*
13: gb_bal:*
14: gb_bal:*
15: gb_bal:*
16: gb_bal:*
17: gb_bal:*
18: gb_bal:*
19: gb_bal:*
20: gb_bal:*
21: gb_bal:*
22: gb_bal:*
23: gb_bal:*
24: gb_bal:*
25: gb_bal:*
26: gb_bal:*
27: gb_bal:*
28: gb_bal:*
29: gb_bal:*
30: gb_bal:*
31: gb_bal:*
32: gb_bal:*
33: gb_bal:*
34: gb_bal:*
35: gb_bal:*
36: gb_bal:*
37: gb_bal:*
38: gb_bal:*
39: gb_bal:*
40: gb_bal:*
41: gb_bal:*
42: gb_bal:*
43: gb_bal:*

44: em_hcg2:*
45: em_hcg3:*
46: em_hcg4:*
47: em_hcg5:*
48: em_hcg6:*
49: em_hcg7:*
50: em_hcg8:*
51: em_hcg9:*
52: em_hcg10:*
53: em_hcg11:*
54: em_hcg12:*
55: em_hcg13:*
56: em_hcg14:*
57: em_hcg15:*
58: em_hcg16:*
59: em_hcg17:*
60: em_hcg18:*
61: em_hcg19:*
62: em_hcg20:*
63: em_hcg21:*
64: em_hcg22:*
65: em_hcg23:*
66: em_hcg24:*
67: em_hcg25:*
68: em_hcg26:*
69: em_hcg27:*
70: em_hcg28:*
71: em_hcg29:*
72: em_hcg30:*
73: em_hcg31:*
74: em_hcg32:*
75: em_hcg33:*
76: em_hcg34:*
77: em_hcg35:*
78: em_hcg36:*
79: em_hcg37:*
80: em_hcg38:*
81: em_hcg39:*
82: em_hcg40:*
83: em_hcg41:*
84: em_hcg42:*
85: em_hcg43:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 1471 | 97.0 | 1505 | 10 | AF137386 Homo sapi |
| 2 | 1193.2 | 78.7 | 141132 | 77 | HS192P9 Human DNA |
| 3 | 934 | 61.6 | 221262 | 30 | AC009090 Homo sapi |
| 4 | 428.8 | 28.3 | 1475 | 11 | RNPPLIPIN |
| 5 | 428.8 | 28.3 | 1622 | 81 | AS1771 |
| 6 | 416 | 27.4 | 1250 | 11 | RNU13617 |
| 7 | 272 | 17.9 | 272 | 79 | G06198 |
| 8 | 124.8 | 8.2 | 853 | 81 | AR062277 |
| 9 | 78.4 | 5.2 | 168747 | 59 | AC078781 |
| 10 | 78.4 | 5.2 | 179017 | 39 | AC012520 |
| 11 | 78.4 | 5.2 | 195115 | 55 | AC046140 |
| 12 | 74 | 4.9 | 1068 | 79 | CNS01E0R |
| 13 | 74 | 4.9 | 2337 | 78 | HSMB00876 |
| 14 | 73.8 | 4.9 | 737 | 56 | TCH2A |
| 15 | 73.8 | 4.9 | 1288 | 56 | TBBDPM5YN |
| 16 | 73.6 | 4.9 | 3916 | 6 | AF097728 |
| 17 | 73.6 | 4.9 | 67554 | 66 | AC083962 |
| 18 | 73.2 | 4.8 | 1582 | 81 | A77033 |
| 19 | 73.2 | 4.8 | 1582 | 81 | A77033 |
| 20 | 73.2 | 4.8 | 1887 | 78 | HSGLNACT |
| 21 | 73.2 | 4.8 | 2896 | 4 | AF061981 |


```

|||||
Db 931 GGGGATCTGAGGCTGTCTCTGCTTGTCTTATGAGAGACTTCAAGCTCCAAAGCTGGG 990
|||
Qy 963 ccaacccctttacccagcactaaatgacatacaagaagactccagactgagccccagac 1022
|||||
Db 991 CCCACCTTCTTCACACACACTAAATGACATPAACAGAGACCTCCACACTGAGCCCCAGAC 1050
|||
Qy 1023 ccgcgcagataagcactaaccaagcaacagctagacacttagcttctggtccagagagc 1082
|||||
Db 1051 CCGCCGATAGTAAAGCTTAACACACACAGTACACCTTAGTCTTGTTCAGAGAGAGC 1110
|||
Qy 1083 tgcgaagctgctgtaaacacactctcctctttaaaccacgcttcaaccaactctccct 1142
|||||
Db 1111 TGAGCAGACTGTGTAACACACTCTCTCTTAAACACCGTTCACACCAACCTCTCCCT 1170
|||
Qy 1143 ggcgcgaactgtataaaagtgggttgattgctgacagcaatgcttccctccctgacatt 1202
|||||
Db 1171 GGAGCCAACTGTAAAAAGGGGTGATGCTGACACATGATGCTTCCCTGCAATTT 1230
|||
Qy 1203 cagacatacagctactgtaagaagcaaatcagctttaaagtattctcagctgtaaaagcc 1262
|||||
Db 1231 CAGACATACAGTACTGAAAGCAATACATTTTAAGTAGTCTCTGAGTCTGAAAGGCC 1290
|||
Qy 1263 tgcgaagcttctcctccttcccaagcctcctctctgtaatactcccttggcggaagct 1322
|||||
Db 1291 TGTCAGAGTTTCCCTTCCCAAGCCCTCTCTGTATACCTCCCTTGGGCGAAGCT 1350
|||
Qy 1323 aacatggtgctcccccagccttctgtaagcagacatgtagcgaaggaaggagagaa 1382
|||||
Db 1351 AACATCGTCTCCCTCCGACCTTGTGACATGACATGAGCAGCAAGAGGAGAGGAA 1410
|||
Qy 1383 gcaagagctgctgctgagctgctgctgctgctgctgctgctgctgctgctgctgctgct 1442
|||||
Db 1411 GCAAGGCTTCCCTGCGGAGATTTGTCTATGCTGTGAGTGAGTCTGATTTATTTTAAAT 1470
|||
Qy 1443 aaaaataaagatgagagaataaaaaaaa 1473
|||||
Db 1471 AAAAATTAAGATGAGAGAAATTAATAAAAAA 1501
|||||

RESULT 2
LOCUS HS192P9 141132 bp DNA PRI 04-FEB-2000
DEFINITION Human DNA sequence from clone RP1-192P9 on chromosome Xp11.23-11.4
Contains a pseudogene similar to rat Plasmolipin, ESTs and GSSs,
complete sequence.
ACCESSION AL020989
VERSION AL020989.2 GI:4539519
KEYWORDS HTG; Plasmolipin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 141132)
AUTHORS Heath, P.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT
On Mar 29, 1999 this sequence version replaced gi:4034132.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
During sequence assembly data is compared over overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession

```

numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X
Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/ChrX>
RP1-192P9 is from the library RPCT-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/>
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP1-192P9.

FEATURES

```

source
1..141132
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="p11.23-11.4"
/clone="RP1-192P9"
/clone_11b="RPCT-1"
1..3583
/note="LIM1 repeat: matches 434..4056 of consensus"
3551..4648
/note="LIP5 repeat: matches 4328..5453 of consensus"
5364..5741
/note="LIM2 repeat: matches 5452..5829 of consensus"
6053..6531
/note="LIM2 repeat: matches 5829..6308 of consensus"
9556..9628
/note="MTL1-internal repeat: matches 1570..1651 of
consensus"
9627..9700
/note="LIM1 repeat: matches 5442..5514 of consensus"
9703..11332
/note="MTL1-internal repeat: matches 3..1561 of
consensus"
14098..14616
/note="MER1A repeat: matches 1..527 of consensus"
14617..14888
/note="LIMB7 repeat: matches 5902..6169 of consensus"
15231..16025
/note="LIMB3A repeat: matches 5313..6159 of consensus"
16066..16404
/note="LIMB3A repeat: matches 4885..5213 of consensus"
19917..20048
/note="LIM8 repeat: matches 271..402 of consensus"
20049..20575
/note="MT2CA repeat: matches 1..501 of consensus"
20576..20907
/note="LIM8 repeat: matches 402..734 of consensus"
20908..21045
/note="LIM8 repeat: matches 6146..6283 of consensus"
22973..23705
/note="LIPB3 repeat: matches 5414..6147 of consensus"
25117..30634
/note="LIPB3 repeat: matches 1537..4181 of consensus"
30635..30885
/note="LIP6 repeat: matches 5893..6143 of consensus"
30886..32124
/note="LIPB3 repeat: matches 4181..5403 of consensus"
34239..35400
/note="LIPB2 repeat: matches 4983..6154 of consensus"
41121..41996
/note="THE1C-internal repeat: matches 1..928 of consensus"
44380..44610
/note="Char1e1 repeat: matches 1..239 of consensus"
44598..44731
/note="Char1e1 repeat: matches 2623..2761 of consensus"
48313..48597
/note="LIP repeat: matches 4181..4465 of consensus"
49218..49531
/note="MER45B repeat: matches 1..322 of consensus"

```

| | | |
|---------------|--------------|--|
| repeat_region | 53820-.54308 | /note="L1PA15 repeat: matches 5661. 6155 of consensus |
| repeat_region | 54310-.55019 | /note="L1PA15 repeat: matches 5601. 6289 of consensus" |
| repeat_region | 55889-.56083 | /note="L1MA7 repeat: matches 5601. 6289 of consensus" |
| repeat_region | 56389-.56432 | /note="MER20 repeat: matches 2. 171 of consensus" |
| repeat_region | 55591-.60059 | /note="MER20 repeat: matches 171. 218 of consensus" |
| repeat_region | 60435-.60594 | /note="MLPIC repeat: matches 1. 466 of consensus" |
| repeat_region | 61509-.62483 | /note="L1MD repeat: matches -17. 160 of consensus" |
| repeat_region | 62544-.65473 | /note="L1MD repeat: matches 973. 1957 of consensus" |
| repeat_region | 65756-.67064 | /note="L1M1 repeat: matches -1389. 1642 of consensus" |
| repeat_region | 67081-.67505 | /note="L1PA7 repeat: matches 4826. 6145 of consensus" |
| repeat_region | 69070-.69210 | /note="L1MD1 repeat: matches 5816. 6224 of consensus" |
| repeat_region | 69493-.69585 | /note="L1R37B repeat: matches 348. 468 of consensus" |
| repeat_region | 69493-.69585 | /note="L1R37B repeat: matches 250. 348 of consensus" |
| repeat_region | 69657-.69748 | /note="L1R37B repeat: matches 126. 198 of consensus" |
| repeat_region | 69925-.70037 | /note="L1R37B repeat: matches 7. 126 of consensus" |
| repeat_region | 70339-.70496 | /note="MER58A repeat: matches 9. 174 of consensus" |
| repeat_region | 72368-.72640 | /note="L1PB1 repeat: matches 5869. 6155 of consensus" |
| repeat_region | 72654-.72968 | /note="MER74B repeat: matches 62. 372 of consensus" |
| repeat_region | 73089-.73136 | /note="24 copies 2 mer tg 97% conserved" |
| repeat_region | 75569-.75582 | /note="12 copies 2 mer ac 100% conserved" |
| repeat_region | 75878-.78882 | /note="L1M1 repeat: matches -1388. 1700 of consensus" |
| repeat_region | 80486-.80644 | /note="L1M1 repeat: matches 1721. 1870 of consensus" |
| repeat_region | 80692-.80838 | /note="L1MA7 repeat: matches 6138. 6282 of consensus" |
| repeat_region | 80915-.81137 | /note="match: GSS: Em:AQ190317" |
| repeat_region | 81138-.83982 | /note="L1M1 repeat: matches 1358. 4230 of consensus" |
| repeat_region | 85760-.85913 | /note="FRAM repeat: matches 1. 154 of consensus" |
| repeat_region | 86852-.87091 | /note="L1PB3 repeat: matches 5909. 6150 of consensus" |
| repeat_region | 89773-.89895 | /note="L1PA15 repeat: matches 6029. 6157 of consensus" |
| repeat_region | 89896-.90141 | /note="ALuSc repeat: matches 1. 235 of consensus" |
| repeat_region | 90143-.90393 | /note="L1P repeat: matches 3337. 3576 of consensus" |
| repeat_region | 9424-.93822 | /note="L1P repeat: matches 4834. 5232 of consensus" |
| repeat_region | 93823-.93958 | /note="L1P repeat: matches 5226. 5362 of consensus" |
| repeat_region | 93966-.94020 | /note="L1PA11 repeat: matches 6110. 6164 of consensus" |
| repeat_region | 94334-.95223 | /note="L1PA15 repeat: matches 5293. 6156 of consensus" |
| repeat_region | 95584-.95680 | /note="MERAD repeat: matches 362. 458 of consensus" |
| repeat_region | 95683-.95901 | /note="MERAD repeat: matches 5. 234 of consensus" |
| repeat_region | 97432-.97616 | /note="MLTIE repeat: matches 343. 510 of consensus" |
| repeat_region | 97923-.97956 | |


```

/dev_stage="adult"
/tissue_type="kidney"
64. .537
CDS

```

| BASE COUNT | 243 | 360 | 332 | 315 |
|------------|-----|-----|-----|-----|
| ORIGIN | a | c | g | t |

| | | | | |
|-----------------------|----------------|---------------|----------|-------------|
| Query Match | 27.4% | Score 416 | DB 11 | Length 1250 |
| Best Local Similarity | 83.3% | Pred. No. 1 | 7e-61 | |
| Matches 473 | Conservative 0 | Mismatches 95 | Indels 0 | Gaps 0 |

QY 113 ccgtcgaagttagcacgcgcgaaccagaagtccttcgcgcagggcgccgaagccttcgtgtcgc 172
|||||
Db 1 CCCTCGAAAGTGAGCACCGACGACCGAGCCCGCGCAGGGCGTGGAGCAATCTGTCTCT 60

```
QY      173 ggcgtcgcgcccgacctggcttcgtcgtcccccctccgaggcgactcaigtctgcag    232
        ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
db       61 GCATGCCAGCCGACCTGGCGTTGTTCGCCTCCGCCCTTGGGGTGCTCGCGCTTTCAG    120
```

233 ctgttgcctgggagctctgtgtgtgggcgtatctgcgacacccgtacacacgtatccg 292
 |||||
 121 ctgctgctgggggctgctggtgtgggaccttgattgctgacaccccatcactctgattcct 180

Dy 923 gccatcgcgagtgatgttcctgcgcgtcttccctcgcgtggacaacgcctcttc 352
| | | | | | | | | | | | | | | | | | | | | |
Db 181 GCCACGGCGTATGTTGTAGCTCTTCCTCTGGCTGGTAACAATTGCTCTTC 240

Dy 353 aaactctacctgtttcagctgcacatgaagtgttacatggttccctggccaactgytgta 412
+
Db 241 ATCATCTACCTGTTTCACATGCACATGAAGTGTATATGGTGCCCTGGCCGTTGTGTTTA 300

413 atgactcttaacatcacgagccacgcgtctctacatcacgcctcatgcctgtctgcg 472
 11 1111 11111111 1111111111 111111 111
 301 ctgctcttcttcttggctctgccacccgctctatattacacgccctttgtgcctgtgcggcg 360

Dy 473 gcaatgactgaacctccctgaggggaacccgcgttaataaccagcgcggcgctgcccg 532
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 GCGGTGATCTGCACATCCCTGAAGGGCTCCCAGCACGGTAGCAACAAGCGTTCGGCTGCCCTCT 420

```

QY      533  ttctctgcgttcttgatgatcgcctatgagatgagcctcttcagctacacgagcc 592
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      421  ttctttgcttgccttgatgatcctacggcactgagtgctttcttcagcttccacgccc 480

```

Dy 593 tggcgagsgagttagcgaacatgcygcaccacagtcagatgctgcygctatgtccaaacc 652
|||||
Db 481 TGCGAGGAGTGGGCACCAACGCCGGCCACGAGTCAGATGGCTGGGGGCTTACTTTAAGCC 540

| | | | |
|----|-----|------------------------------|-----|
| Oy | 653 | acgtgtgcaacgccccctctgggagctg | 680 |
| | | | |
| Db | 541 | AGCTATGCTGTGACCTAAGCCACGGCTG | 568 |

| RESULT | 7 |
|----------|-------------|
| G06198/c | |
| LOCUS | G06198 |
| | 272 bp |
| DNA | |
| STS | 19-OCT-1995 |

| | |
|-----------------------|----------------------|
| DEFINITION | human STS WI-62/5. |
| ACCESSION | G06198 |
| VERSION | G06198.1 GI:859443 |
| KEYWORDS | |
| STS sequence: primer: | sequence tagged site |

| SOURCE | ORGANISM |
|---|--|
| human SVS derived from sequences in dbEST and the Unigene collection. | Homo sapiens |
| | Eukaryotae: mitochondrial eukaryotes. Metazoa: Chordata: |

Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Hominidae; Homo.

| REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
|-----------|------------------|------------|--|--------------------|
| 1 | (bases 1 to 272) | Hudson, T. | Whitehead Institute/MIT Center for Genome Research: Physically Mapped ESTs | Unpublished (1995) |

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: AACAGTCACCACCACACCACA
Primer B: TTCAGACATACCAGTTACTGAAGC
Sns size: 228
PCR Profile:

Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Amplified: 10 µl

Template: 10 ng
 Primer: each 5 pM
 dNTPs: each 4 nM
 Taq Polymerase: 0.025 units/ μ l
 Total Vol: 20 μ l

buffer:
MgCl₂: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 9.3

Prepared with primer pairs derived from Z38289 -- dbEST.
Location/Qualifiers

| | |
|-------------|--------------------------|
| 5MS | /organism="Homo sapiens" |
| primer_bind | 37. .264 |
| primer_bind | 37. .56 |
| BASE COUNT | complement(240. .264) |
| ORIGIN | 79 a 64 c 61 g 68 t |

| | | | | |
|-----------------------|----------------|-------------------|----------|------------|
| Query Match | 17.9% | Score 272 | DB 79 | Length 272 |
| Best Local Similarity | 100.0% | Pred. No. 6.2e-37 | | |
| Matches 272 | Conservative 0 | Mismatches 0 | Indels 0 | Gaps 0 |

DQ 1193 ccctgcatttcagacatccagttaactgnaagcaaatcagttttaagtgtattctcagtg 1252
|||||
|||
Db 272 ccctgcatttcagacatccagtttaactgnaagcaaatcagttttaagtgtattctcagtg 213

QY 1253 ctgaagagcctgtccagggttccttccttcctcccaagcctctctcytaatctccctt 1312
|||||
Dd 212 ctgaaaaagcctgtccagggttccttccttcctcccaagcctctctctgtaactcccttt 153
|||||

| | | | |
|----|--|---|------|
| OY | | 1313 gggcgaagctaacaatcggtgctccccaccattgcctgactaggacatatygtagcaagaag | 1372 |
| Dδ | 152 GGGCGAAGCTAACATCGGTCGCCCTCCCGACCTTGTGTGACTAGGCACATGGCAGCAAGAAG | G93 | |

Qy 1373 agggagggaaagcaagccttgcccgagatgtcatagtgtgtgtgacgtttcat 1382
|||||
Db 92 AGGAGGGAAGCAAGCCTTGCCGCGGAGTGTATGTGCTGTGCTGACGCTTTAT 33

| | | |
|----|------|--------------------------------|
| Oy | 1433 | ttttttaaataaagatgagagaatt |
| | | |
| Db | 32 | TTTTTTTAAATAAATTAAGATGACAGAATT |

/product="hypothetical protein"
/protein_id="CAB53680.1"
/db_xref="GI:5817155"
/translation="LERRRELSRLGDKKPEQAVEKYTHIIGLSKKDGLVPMFINTH
SGLETHGVFTLGARADSYEYLILKOWOGGKOETQLLEDYVEAIEGYRTLLRSEH
SKLPFVELAHGRFSAKMDHVCEFLPGTLAGVHGHPASHEMLAEOLMENCYOMNR
METLSPEIVHFENLYPOGRDRDEVKPADRNLLRPFEVESLFYRYTGDRKODMGW
WEIIQSFRERFSTCCPRPMSRPPGHGMAVGILRIAPLIVAVTWBREGARSLG
VATIQLGSPGPIPTPELLHGCTOSTWRKMALIFITSGLSNREGICCPSTIPPV"
polyA_site
BASE COUNT 494 a 669 c 732 g 442 t
ORIGIN

Query Match 4.9% Score 74; DB 78; Length 2337;
Best Local Similarity 60.3%; Pred. No. 0.0034;
Matches 140; Conservativity 0; Mismatches 90; Indels 2; Gaps 1;

Oy 1285 ccaagccctcctcttaataactccttcgttgagggaagcctaactgctgcctcccgacct 1344
||| || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2059 CCAGGGCGTCGCCTGTGTTACAAAGTCGACTCAGGGATCCTCTGCCGCCCG--CA 2116

Oy 1345 tcgtactagagcacatggagacgaagagaggaagaagcaagcgcttgcgcggagatt 1404
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2117 GGGGCGTTGGAAGCGCTGAGCGCAAGTCGCTACTCATCGGCCCCCTCCAATGANAAG 2176

Oy 1405 gtcatgttgttgttgtgactgttttatltttttaataaaataagaatgagaagt 1464
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2177 GCTTTTCGTGCGATTAAGTGTGATTGCTCTAAAAAAAATAAAAAAAAAAAAAA 2236

Oy 1465 aa 1516
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2237 AA 2288

RESULT 14
LOCUS TCH2A 737 bp mRNA INV 18-JUN-1998
DEFINITION Trypanosoma cruzi cDNA for histone H2A.
ACCESSION Y13985
VERSION Y13985.1 GI:2222801
KEYWORDS histone H2A.
SOURCE Trypanosoma cruzi.
ORGANISM Trypanosoma cruzi.
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE Lopez,M.C.
1 (bases 1 to 737)
DIRECT Submission
Submitted (23-JUN-1997) Lopez M.C., Molecular Biology, Instituto de
Parasitologia y Biomedicina 'Lopez Neyra', C.S.I.C., Ventanilla, 11
18001 Granada SPAIN
2 (bases 1 to 737)
Matanon,C., Puerta,C., Alonso,C. and Lopez,M.C.
Control mechanisms of the H2A genes expression in Trypanosoma cruzi
Mol. Biochem. Parasitol. 92 (2), 313-324 (1998)
98319417
FEATURES
source location/Qualifiers
1..737
/organism="Trypanosoma cruzi"
/strain="1"
/db_xref="taxon:5693"
/dev_stage="Amastigote"
98..643
/gene="h2a"
98..505
/gene="h2a"
/codon_start=1
/evidence="experimental"
/product="Histone H2A"
/protein_id="CAA74318.1"
/db_xref="GI:2222802"
/db_xref="SPTREMBL:O02617"

| | | |
|-----------------------|---|---|
| | polyA_site | /translation="MAPPKOAKRASKKHGGGRSKAGCLIPVGVSGLRRGOYAA IGASGAYVMAAVILEYTALLETLELSVKASQOAKPKRLPTPTVTILVAVRHDDDLGMLL NVTLSRGGVIFSLNKAAVAKKHKSSKKARATPSA" 643 /gene="h2a" |
| BASE COUNT | 219 a 179 c 208 g 131 t | /evidence=experimental |
| ORIGIN | | |
| Query Match | 4.9%; Score 73.8; DB 56; Length 737; | |
| Best Local Similarity | 72.2%; Pred. No. 0.0038; | |
| Matches | 96; Conservative 0; Mismatches 37; Indels 0; Gaps 0; | |
| Oy | 1384 caaagcctgcctgacgagtgtcatgttggttggtagtactttattttaata 1443 | |
| Dd | 588 CCACCCGTGACCCCTTACAATAGCATGTGTTGTGTGTTCCTTTGCTTTCTAANA 647 | |
| Oy | 1444 aaataagaatgaggaatttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1503 | |
| Dd | 648 AA 707 | |
| Oy | 1504 aaaaaaaaaaaa 1516 | |
| Dd | 708 AAAAAAAAAAAAAA 720 | |
| RESULT 15 | | |
| TBBDPM SYN | 1288 bp mRNA INV 01-NOV-1996 | |
| LOCUS | T.brucei brucei mRNA for dolichyl-phosphate-mannose synthase. | |
| DEFINITION | Z54162 | |
| ACCESSION | Z54162.1 GI:1150647 | |
| KEYWORDS | dolichyl-phosphate-mannose synthase. | |
| SOURCE | Trypanosoma brucei brucei. | |
| ORGANISM | Trypanosoma brucei brucei Eukaryota; Elenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma. 1 (bases 1 to 1288) Mazhari-Tabrizi,R., Eckert,V., Blank,M., Muller,R., Mumberg,D., Funk,M. and Schwarz,R.T. Cloning and functional expression of glycosyltransferases from parasitic protozoans by heterologous complementation in yeast: the dolichol phosphate mannose synthase from Trypanosoma brucei brucei Blochem. J. 316 (Pt 3), 853-858 (1996) 96265052 2 (sites) 2 (bases 1 to 1288) Mazhari,R. Direct Submission Submitted (28-AUG-1995) Mazhari R., Zentrum fuer Hygiene und Medizinische Mikrobiologie, Parasitologie, Robert-Koch-Str. 17, Marburg, Hessen, Germany, 35037 Location/Qualifiers 1. .1288 /organism="Trypanosoma brucei brucei" /strain="118" /db_xref="taxon:5702" 200. 204 214. :300 214. :1017 /EC_number="2.4.1.83" /codon_start=1 /product="dolichyl-phosphate-mannose synthase" /protein_id="CA90875.1" /db_xref="GI:1150648" /db_xref="SptREMBL:Q26732" /translation="MAVKYSIIPAYKECGNLISLQAQVDRLADDFESKNVEWIVD DNSDGSVEVEKYKRNEGYGVRIVRTDRGLSSAVIHGISVSKSGFLIVNDADLOHP PKTYCILLRALEKRGVEVCSTRYGAGVEIDKDWPLHRRTISMGARLLARPISDP MSGFGFSRMSEFNGSCRVNPDIQTALTELPVCAVARYEVGFNFPAARTVGSKLTL GKVITNIHLKLITYFYTGALTGSTFAVAGLGFTYTFPFFTHLSLA" | |
| JOURNAL MEDLINE | | |
| REFERENCE | | |
| AUTHORS | | |
| TITLE | | |
| JOURNAL | | |
| FEATURES | | |
| source | | |
| RBS | | |
| sig.peptide | | |
| CDS | | |
| BASE COUNT | 388 a 243 c 313 g 344 t | |

XX 02-SEP-1999; 99WO-US20468.
 PF 02-SEP-1998; 98US-0145815.
 XX 12-NOV-1998; 98US-0191283.
 PR 09-DEC-1998; 98US-0208821.
 PR 26-JAN-1999; 99US-0237506.
 PR 10-FEB-1999; 99US-0247891.
 XX (INCYTE) INCYTE PHARM INC.
 XX Au-Young J, Bandman O, Tang Y, Reddy R, Hillman JL, Yue H;
 PI Lal P, Corley NC, Guegler KJ, Gorgone G, Baughn MR, Azimzai Y;
 XX WPI: 2000-256643/22.
 DR P-PSDB: Y70458.
 DR Novel human membrane channel protein and polynucleotide useful for
 PT diagnosing and treating cell proliferative, inflammatory, secretory,
 PT osmoregulatory, muscular, cardiovascular and neurological disorders
 XX
 PS Claim 9; Page 123; 140pp; English.
 CC The present sequence is a cDNA identified in Incyte clone 1375415
 CC derived from LUNGNOT10 cDNA library. It encodes human membrane channel
 CC protein-8 (MECHP-8), which is expressed in nervous and gastrointestinal
 CC tissues. Anti-MECHP antibodies can be used as therapeutic antagonists and
 CC reagents for diagnosis and monitoring diseases. MECHP cDNA can be used
 CC for diagnosis of MECHP-related diseases and gene mapping. MECHP can be
 CC used for treatment of cell proliferative disorders such as bursitis and
 CC atherosclerosis, cancers like lymphoma and sarcoma, inflammatory
 CC disorders like AIDS and Addison's disease, transport/secretory disorders
 CC like cystic fibrosis and diabetes mellitus, osmoregulatory disorders like
 CC diarrhoea and renal failure, muscular disorders like myocarditis and
 CC Duchenne's muscular dystrophy, cardiovascular disorders like hypertension
 CC and vasculitis, congenital lung anomalies like bronchitis and asthma and
 CC neurological disorders like Alzheimer's disease, Parkinson's disease and
 CC Huntington's disease.
 XX
 XX Sequence 884 BP; 146 A; 280 C; 282 G; 173 T; 3 other;
 SO
 Query Match 55.6%; Score 842.4; DB 21; Length 884;
 Best Local Similarity 99.5%; Pred. No. 1e-133;
 Matches 854; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
 QY 3 tcggagagctgggggaaagcgcgcgagcgagccacccacgctggagggcgcgac 62
 DB 22 tcggagagctgggggaaagcgcgcgagcgagccacccacgctggagggcgcgac 81
 QY 63 ggcggccgtagcagcctcgagagcgagcgagccgcatggcgagctccgctcgaaag 122
 DB 82 ggcggccgtagcagcctcgagagcgagcgagccgcatggcgagctccgctcgaaag 141
 QY 123 ttaagcagcgagcagcagcagctctcgagcgagcgagccgagctggctcgagctggcc 182
 DB 142 ttaagcagcgagcagcagcagctctcgagcgagcgagccgagctggctcgagctggcc 201
 QY 183 cggagcctgggcttctgtgctcccgctcgagcgagctatgctgtgagctgtgtgtg 242
 DB 202 cggagcctgggcttctgtgctcccgctcgagcgagctatgctgtgagctgtgtgtg 261
 QY 243 ggcgcctgtgtgggagctgtatcgagcagccgctacacccgtatccgagctatgtgt 302
 DB 262 ggcgcctgtgtgggagctgtatcgagcagccgctacacccgtatccgagctatgtgt 321
 QY 303 ggcgtgatgtctgtctgtctctctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 362
 DB 322 ggcgtgatgtctgtctgtctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 381
 QY 363 tttttcagctcagatgaagt 422
 DB 382 tttttcagctcagatgaagt 441

QY 423 acatcagcgccacgcttctacatcacccgcttccatcgctgtctgtcggaattgacc 482
 DB 442 acatcagcgccacgcttctacatcacccgcttccatcgctgtctgtcggaattgacc 501
 QY 483 tgacatccctgaagggaacccgagcctataacacgagcgagctgtctgtctgtct 542
 DB 502 tgacatccctgaagggaacccgagcctataacacgagcgagctgtctgtctgtct 561
 QY 543 gtttggtatgatgagcctatgagtgagtgctctcttccatgacccagcctggagag 602
 DB 562 gtttggtatgatgagcctatgagtgagtgctctcttccatgacccagcctggagag 621
 QY 603 tagcagcagatgcgagcagcagcagcagcagcagcagcagcagcagcagcagcag 662
 DB 622 tagcagcagatgcgagcagcagcagcagcagcagcagcagcagcagcagcagcag 681
 QY 663 cggcccccctgtgggtgtgaagcgcgagctgtgtgtgtgtgtgtgtgtgtgtgtgt 722
 DB 682 cggcccccctgtgggtgtgaagcgcgagctgtgtgtgtgtgtgtgtgtgtgtgtgt 741
 QY 723 gaagctgggagcctgt 781
 DB 742 gaagctgggagcctgt 801
 QY 782 ccgtcagacataagctctcagcagcagcagcagcagcagcagcagcagcagcagcag 841
 DB 802 ccgtcagacataagctctcagcagcagcagcagcagcagcagcagcagcagcagcag 861
 QY 842 ctgcagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 859
 DB 862 ctgcagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 879
 RESULT 2
 T34592
 ID T34592 standard; cDNA; 1622 BP.
 XX T34592; .
 AC T34592; .
 XX 01-JAN-1997 (first entry)
 DT
 XX NTII-11 nerve protein coding sequence.
 DE
 XX Probe; central nervous system; peripheral nervous system; CNS; PNS;
 KW axon; neurone; neuronal cells; glial cells; Schwann cells; trauma;
 KW pathology; regeneration; MS2; metalloproteinase; ds.
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 362..907 /tag= a
 FT /product= NTII-11 protein expressed in damaged or
 FT regenerating nerve cells.
 XX W09617865-A2.
 XX PD 13-JUN-1996.
 XX PF 05-DEC-1995; 95WO-EP04777.
 XX PR 27-JAN-1995; 95DE-4002525.
 PR 05-DEC-1994; 94DE-4443159.
 XX (BOE) BOEHRINGER MANNHEIM GMBH.
 PI Gillen C, Gleichmann M, Mueller H;
 XX WPI: 1996-287114/29.
 DR P-PSDB: R99799.
 XX Nucleic acids associated with damaged or regenerating nerve cells

Db 181 gcaagctgtgaaccacactctccctccttaacaacngtllcaaacancccttncntgga 240
 Qy 116 gccaccgcttaaaagtggtgattgctgacagcattgcttccctccctcattcag 1205
 Db 241 gccacacngtaaaagtggtgattgctgacagcattgcttccctccctcattcag 300
 Qy 1206 acatacagctt 1216
 Db 301 acatacagctt 311
 RESULT 4
 ID X10969/c
 XX X10969 standard; DNA; 228 BP.
 AC X10969;
 DT 30-MAR-1999 (first entry)
 DE Human biallelic polymorphic DNA fragment WI-6275.
 KM Polymorphism; biallelic; human; forensic; paternity testing; disease;
 KW detection; phenotypic typing; characteristic; infection; hereditary;
 KW autoimmune disease; cancer; inflammation; drug; therapy; medication;
 KW treatment; marker; ss.
 XX Homo sapiens.
 OS
 XX W09820165-A2.
 PN 14-MAY-1998.
 PD 05-NOV-1997; 97WO-US20313.
 PF 06-NOV-1996; 96US-0030455.
 PR (MHED) WHITEHEAD INST BIOMEDICAL RES.
 XX Hudson T, Lander ES, Wang D;
 PI WPI: 1998-286974/25.
 DR
 XX New isolated nucleic acid segments from the human genome - used for
 PT determining polymorphic forms for use in e.g. forensics, paternity
 PT testing or phenotypic typing for disease
 PS Claim 1; Page 101; 310pp; English.
 XX X10269-X12937 are human DNA fragments which contain biallelic polymorphic
 CC markers which have been isolated using the primers represented in
 CC X09121-X10268. The base occupying the polymorphic site is indicated by
 CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
 CC methods for determining polymorphic forms in an individual for use in
 CC e.g. forensics, paternity testing or for phenotypic typing for diseases
 CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
 CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
 CC hypercholesterolemia, polycystic kidney disease, hereditary
 CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
 CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
 CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
 CC autoimmune diseases, inflammation, cancer, diseases of the nervous
 CC system, infection by pathogenic microorganisms, and characteristics such
 CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
 CC endurance, fertility, and susceptibility or receptivity to particular
 CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
 CC segments can also be used to produce medicaments for the treatment or
 CC prophylaxis of such diseases.
 SQ Sequence 228 BP; 62 A; 59 C; 56 G; 50 T; 1 other;

Query Match 15.0%; Score 227.6; DB 19; Length 228;

Best Local Similarity 99.6%; Pred. No. 2.4e-30;
 Matches 227; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Qy 1201 ttccagctaccagcttactgaagaacaaatcagtttaagtgtctcgaagctgaag 1260
 Db 228 ttccagctaccagcttactgaagaacaaatcagtttaagtgtctcgaagctgaag 169
 Qy 1261 cctgtccaggttctcctccttcccaagcctctctctgtaatactcccttggcgaag 1320
 Db 168 cctgtccaggttctcctccttcccaagcctctctctgtaatactcccttggcgaag 109
 Qy 1321 ctacactgggtgctctcccgacctgtgctgactaggacacatggagcgaagaggagg 1380
 Db 108 ctacactgggtgctctcccgacctgtgctgactaggacacatggagcgaagaggagg 49
 Qy 1381 aagcaagccttgcctgctgaggtgtcattggtgtgtgtgactgtc 1428
 Db 48 aagcaagccttgcctgctgaggtgtcattggtgtgtgtgactgtc 1
 RESULT 5
 ID Z36249
 XX Z36249 standard; cDNA; 1152 BP.
 AC Z36249;
 DT 22-FEB-2000 (first entry)
 DE cDNA encoding a bone marrow secreted protein designated BMS61.
 XX
 KM Bone marrow secreted protein; bone marrow stromal cell; cytokine;
 KW cell proliferation; cell differentiation; hematopoiesis; anemia;
 KW myeloid cell deficiency; lymphoid cell deficiency; myeloid cell;
 KW erythroid progenitor cell; colony stimulating factor; granulocyte;
 KW monocyte; macrophage; myelo-suppression; megakaryocyte; platelet;
 KW platelet disorder; thrombocytopenia; hematopoietic stem cell;
 KW stem cell disorder; aplastic anaemia; bone differentiation;
 KW paroxysmal nocturnal hemoglobinuria; bone growth; cartilage; tendon;
 KW ligament; nerve; wound healing; tissue repair; burn; incision; ulcer;
 KW bone fracture; cartilage damage; artificial joint; ss.
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 276..797
 FT /tag=a
 FT /product="bone marrow secreted protein"
 XX
 PN W09933979-A2.
 PD 08-JUL-1999.
 PF 18-DEC-1998; 98WO-US27008.
 PR 30-DEC-1997; 97US-0068958.
 PR 24-SEP-1998; 98US-0101603.
 PR 30-SEP-1998; 98US-0102540.
 PA (CHIR) CHIRON CORP.
 XX Lin H, Cao L;
 PI WPI: 2000-038344/03.
 DR P-PSDB: Y53643.
 DE
 XX New isolated human polynucleotide and secreted proteins can induce
 PT production of other cytokines in certain cell populations -
 PS Claim 11; Page 118-119; 120pp; English.
 CC Z36228-49 encode bone marrow secreted proteins of human bone marrow
 CC stromal cells. The proteins can exhibit cytokine, cell proliferation, or
 CC cell differentiation activity (either inducing or inhibiting). They can


```

XX 19-OCT-1994; 94EP-0402353.
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX Melki J, Munnich A;
XX WPI; 1996-202055/21.
XX New isolated survival motor neuron gene - used to develop prods. for
XX the diagnosis and treatment of motor neuron diseases
XX Claim 4; Fig 2a; 29pp; English.
XX The SMN gene is a chromosome 5-SMA (Spinal Muscular Atrophy)
XX determining gene. Sequences of clone T-BCD541 and clone C-BCD541
XX are provided in T28255 and T28259, respectively.
XX SMN sequences can be used for detecting neuron disorders having
XX symptoms of muscular weakness with or without sensory changes
XX such as amyotrophic lateral sclerosis (ALS), SMA, primary lateral
XX sclerosis (PLS), etc.
XX Sequence 1582 BP; 562 A; 285 C; 335 G; 400 T; 0 other;

Query Match
Best Local Similarity 86.2%; Score 73.2; DB 17; Length 1582;
Matches 81; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1423 acgtttatttttttaataaataagatgagagaatttaaaaaaaaaaaaaa 1482
DB 1473 aaattttaattttttttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1532
QY 1483 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516
DB 1533 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1566

RESULT 9
TID T18831
XX T18831 standard; cDNA; 1582 BP.
XX T18831;
XX 02-OCT-1996 (first entry)
XX DE Human survival motor neuron variant gene cDNA clone C-BCD541.
XX KW Survival motor neuron gene; SMN gene; spinal muscular atrophy;
XX chromosome 5-SMA determining gene; amyotrophic lateral sclerosis;
XX primary lateral sclerosis; arthrogryposis multiplex congenita;
XX diagnosis; gene therapy; C-BCD541; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 34..918
FT /tag- a
FT 1..114
FT /tag- b
FT /label= Exon-1
FT 115..306
FT /tag- c
FT /label= Exon-2
FT 307..507
FT /tag- d
FT /label= Exon-3
FT 508..660
FT /tag- e
FT /label= Exon-4
FT 661..756
FT /tag- f
FT /label= Exon-5

```

```

FT exon 757..867
FT /tag- g
FT /label= Exon-6
FT exon 868..921
FT /tag- h
FT /label= Exon-7
FT variation 873
FT /tag- i
FT /note= "base 873 (c) is c is T-BCD541"
FT exon 922..1483
FT /tag- j
FT /label= Exon-8
FT variation 1155
FT /tag- k
FT /note= "base 1155 (a) is g in T-BCD541"
FT PN EF711833-A2.
FT PD 15-MAY-1996.
FT XX 19-OCT-1995; 95EP-0402335.
FT PR 19-OCT-1994; 94EP-0402353.
FT PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
FT PI Melki J, Munnich A;
FT WPI; 1996-232098/24.
FT XX Human survival motor neuron gene T-BCD541, variant C-BCD541 and
PT murine equi. - useful to develop primers and probes for in vitro
PT detection of motor neuron diseases e.g. spinal muscular atrophy
PT Claim 4; Fig 2; 47pp; English.
XX PS A cDNA sequence (T18831) was detd. for the human survival motor
XX CC neuron (SMN) variant gene C-BCD541. This centromeric gene
XX CC differs from the SMN telomeric gene cDNA clone T-BCD541 (T18828)
XX CC by 2 nucleotides. The C-BCD541 gene is unable to correct in
XX CC vivo a deficiency of the T-BCD541 gene responsible for motor
XX CC neuron diseases of the spinal muscular atrophy type. A variation
XX CC of sequence within intron 7 of the C-BCD541 genomic sequence
XX CC results in an alternatively spliced transcript that lacks exon 7
XX CC and a truncated SMN protein (R94965) lacking the exon 7-encoded
XX CC C-terminal region.
XX SQ Sequence 1582 BP; 562 A; 283 C; 335 G; 402 T; 0 other;

Query Match
Best Local Similarity 86.2%; Score 73.2; DB 17; Length 1582;
Matches 81; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1423 acgtttatttttttaataaataagatgagagaatttaaaaaaaaaaaaaa 1482
DB 1473 aaattttaattttttttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1532
QY 1483 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516
DB 1533 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1566

RESULT 10
V34315
ID V34315 standard; DNA; 1114 BP.
XX V34315;
XX 29-JAN-1999 (first entry)
XX DE Human secreted protein gene 5 clone HEIDY41.
XX

```

KW Human: secreted protein; fusion protein; gene therapy; protein therapy;
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
XX W09840483-A2.
XX
XX 17-SEP-1998.
XX
XX 12-MAR-1998; 98WO-US04858.
XX
XX 19-DEC-1997; 97US-0068368.
XX
XX 14-MAR-1997; 97US-0040710.
XX
XX 14-MAR-1997; 97US-0040762.
XX
XX 30-MAY-1997; 97US-0048100.
XX
XX 30-MAY-1997; 97US-0048189.
XX
XX 30-MAY-1997; 97US-0048357.
XX
XX 30-MAY-1997; 97US-0050934.
XX
XX 06-JUN-1997; 97US-0048970.
XX
XX 05-SEP-1997; 97US-0057765.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ferrie AM, Fischer CL, Gentz RL, Greene JM, Kyaw H;
XX Li H, Li Y, Moore PA, Rosen CA, Ruben SM, Soppet DR;
XX Wei YF, Young PE, Zeng Z;
XX
XX WPI: 1998-520811/44.
XX
XX P-PSDB; W75225.
XX
XX Isolated human poly:nucleotide(s) encoding secretory peptide(s) -
XX used to develop products for the diagnosis and treatment of e.g.
XX PT inflammation, cancers, CNS disorders or immune system disorders
XX
XX PS Claim 1; Page 142-143; 201pp; English.
XX
XX This sequence represents a nucleic acid molecule which encodes a secreted
XX human protein. The gene number, and the clone it is derived from, are
XX detailed in the descriptor line. The gene can be used to generate fusion
XX proteins by linking to the gene to a human immunoglobulin Fc portion
XX (e.g. V34277) for increasing the stability of the fused protein as
XX compared to the human protein only.
XX The invention relates to 28 novel genes and their fragments (nucleic acid
XX sequences: V34286-V34325; amino acid sequences W75196-W75235) which
XX are useful for preventing, treating or ameliorating medical conditions
XX e.g. by protein or gene therapy. Also, pathological conditions can be
XX diagnosed by determining the amount of the new polypeptides in a sample
XX or by determining the presence of mutations in the new polynucleotides.
XX Specific uses are described for each of the 28 polynucleotides, based on
XX which tissues they are most highly expressed in (see V34286 for described
XX uses).
XX
XX Sequence 1114 BP; 371 A; 196 C; 272 G; 272 T; 3 other;
SQ

Query Match 4.8%; Score 72.2; DB 19; Length 1114;
Best Local Similarity 80.8%; Pred. No. 0.00041;
Matches 80; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

OY 1418 tggtagcgtttatttttttaataaataaagatgaggaatttaaaaaaaaaa 1477
DB 1015 kgttggcttttttttttaaaaaaataaaaaaataaaaaaataaaaaa 1074
OY 1478 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516
DB 1075 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1113

RESULT 11
T71322/c
ID T71322 standard; DNA; 10288 BP.
XX
XX T71322;
XX
XX 11-SEP-1997 (first entry)
XX
XX Plasmid pCB51 encoding partial UNC-53 protein used as bait vector.
XX
XX
XX UNC-53; neuronal regeneration; revascularisation; wound healing;
XX neurodegenerative disease; Alzheimer's disease; Huntington's;
XX peripheral neuropathies; metastasis inhibition; cancer;
XX circular; Caenorhabditis elegans; ds.
XX
XX
XX Synthetic.
XX
XX W09638555-A2.
XX
XX 05-DEC-1996.
XX
XX 31-MAY-1996; 96WO-EP02311.
XX
XX 31-MAY-1995; 95GB-0010944.
XX
XX (BOGA/) BOGAERT T.
XX (STRIL/) STRINGHAM E.
XX (VAND/) VANDEKERCKHOVE J.
XX
XX Bogaert T, Stringham E, Vandekerckhove J;
XX
XX WPI: 1997-034369/03.
XX
XX Caenorhabditis elegans UNC-53 protein 8A and 7A variants - useful to
XX promote neuronal regeneration, revascularisation or wound healing.
XX
XX Example 15; Page 139-145; 278pp; English.
XX
XX T71321 and T71322 are plasmids pCB50 and pCB51, respectively. They
XX are used as bait vectors in a yeast two hybrid system. pCB50 encodes
XX the full length UNC-53 protein variant 7A of Caenorhabditis elegans
XX and pCB51 encodes a C-terminal portion of this protein. Both vectors
XX were constructed by cloning the appropriate cDNA fragment into
XX pAS1-CH2 (Clontech). The yeast two hybrid system was set up for
XX identifying proteins that interact with UNC-53. UNC-53 protein 8A and
XX 7A variants and nucleic acids encoding them are useful as medicaments
XX to promote neuronal regeneration, revascularisation or wound healing, or
XX for treatment of chronic neurodegenerative diseases (e.g. Alzheimer's or
XX Huntington's disease) or acute traumatic injuries. Transgenic cells and
XX organisms transfected with UNC-53 cDNA can be used to determine whether
XX a substance is an inhibitor or enhancer of the regulation of cell shape
XX or motility or the direction of cell migration by screening for a
XX phenotypic change in the cell. Inhibitors can be used to alleviate the
XX spread of disease inducing cells or metastasis. Probes derived from the
XX cDNA sequences can be used to identify homologues of the C. elegans
XX unc-53 gene. The UNC-53 protein can be used to identify proteins which
XX are active in the signal transduction pathway that can be used as
XX mentioned above.
XX
XX Sequence 10288 BP; 2909 A; 2218 C; 2242 G; 2918 T; 1 other;
SQ

Query Match 4.8%; Score 72.2; DB 18; Length 10288;
Best Local Similarity 86.0%; Pred. No. 0.00054;
Matches 80; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 1424 ctgtttatttttttaataaataaagatgaggaatttaaaaaaaaaa 1483
DB 8096 CTGTATTAFTTCTATGCAAAAAAAAAAAAAAAAAAAAAAAAAA 8037
OY 1484 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516

Db 8036 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 8004

RESULT 12

ID T71321/C

AC T71321 standard; DNA; 13414 BP.

XX T71321;

DF 11-SEP-1997 (first entry)

DE Plasmid pCB50 encoding UNC-53 protein variant 7A used as bait vector.

XX

XX UNC-53; neuronal regeneration; revascularisation; wound healing;

KW neurodegenerative disease; Alzheimer's disease; Huntington's;

KW peripheral neuropathies; metastasis inhibition; cancer;

KW circular; Caenorhabditis elegans; ds.

XX

OS Synthetic.

XX MO9638555-A2.

PN 05-DEC-1996.

PD

XX 31-MAY-1996; 96WO-EP02311.

PF

XX 31-MAY-1995; 95GB-0010944.

PR

XX (BOGA/) BOGAERT T.

PA (STRI/) STRINGHAM E.

PA (VAND/) VANDERKERCHEVE J.

XX

PI Bogaert T, Stringham E, Vandekerckhove J;

DR WPI; 1997-034369/03.

XX

XX Caenorhabditis elegans UNC-53 protein 8A and 7A variants - useful to

PT promote neuronal regeneration, revascularisation or wound healing.

XX

PS Example 15; Page 132-139; 278pp; English.

XX

XX T71321 and T71322 are plasmids pCB50 and pCB51, respectively. They

CC are used as bait vectors in a yeast two hybrid system. pCB50 encodes

CC the full length UNC-53 protein variant 7A of Caenorhabditis elegans

CC and pCB51 encodes a C-terminal portion of this protein. Both vectors

CC were constructed by cloning the appropriate cDNA fragment into

CC pAS1-CYH2 (Clontech). The yeast two hybrid system was set up for

CC identifying proteins that interact with UNC-53. UNC-53 protein 8A and

CC 7A variants and nucleic acids encoding them are useful as medicaments

CC to promote neuronal regeneration, revascularisation or wound healing, or

CC for treatment of chronic neurodegenerative diseases (e.g. Alzheimer's or

CC Huntington's disease) or acute traumatic injuries. Transgenic cells and

CC organisms transfected with UNC-53 cDNA can be used to determine whether

CC a substance is an inhibitor or enhancer of the regulation of cell shape

CC or motility or the direction of cell migration by screening for a

CC phenotypic change in the cell. Inhibitors can be used to alleviate the

CC spread of disease inducing cells or metastasis. Probes derived from the

CC cDNA sequences can be used to identify homologues of the C. elegans

CC unc-53 gene. The UNC-53 protein can be used to identify proteins which

CC are active in the signal transduction pathway that can be used as

CC mentioned above.

XX

XX Sequence 13414 BP; 3854 A; 3072 C; 2864 G; 3603 T; 1 other;

SO

Query Match 4.8%; Score 72.2; DB 18; Length 13414;

Best Local Similarity 86.0%; Pred. No. 0.00056;

Matches 80; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 1424 cgtgttatttttttaataaataagatgagagaatttaaaaaaaaaaaaaa 1483

DB 11222 CTTGTAATTTCTATGAAAAAATAAAAAAAAAAAAAAAAAAAAAA 11163

OY 1484 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516

DB 11162 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 11130

RESULT 13

ID X37388

AC X37388 standard; cDNA; 773 BP.

XX X37388;

XX

DT 06-JUL-1999 (first entry)

DE Human secreted protein cDNA fragment containing gene 20.

XX

XX Human; secreted protein; prevention; treatment; protein therapy;

KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;

KW developmental abnormality; foetal deficiency; blood disorder; lymphoma;

KW leukemia; immune system disorder; autoimmune disease; hepatic disease;

KW renal disease; inflammation; allergy; asthma; sepsis; diabetes; AIDS;

KW Alzheimer's disease; cognitive disorder; schizophrenia; osteoporosis;

KW arthritis; psoriasis; digestive; endocrine; infection; ss.

XX

XX Homo sapiens.

XX

PN MO9909155-A1.

PD

XX 25-FEB-1999.

PF

XX 18-AUG-1998; 98WO-US17044.

XX

XX 16-JUN-1998; 98US-0092956.

PR 15-JUL-1998; 98US-0092956.

PR 19-AUG-1997; 97US-0056368.

PR 19-AUG-1997; 97US-0056369.

PR 19-AUG-1997; 97US-0056355.

PR 19-AUG-1997; 97US-0056555.

PR 19-AUG-1997; 97US-0056556.

PR 19-AUG-1997; 97US-0056628.

PR 19-AUG-1997; 97US-0056629.

PR 19-AUG-1997; 97US-0056726.

PR 19-AUG-1997; 97US-0056728.

XX

XX (HMAN-) HUMAN GENOME SCI INC.

XX

XX Brewer LA, Duan R, Ebner R, Endress GA, Feng P;

PI Florence C, Florence KA, Komatsoulis GA, Lafleur DW;

PI Moore PA, Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR;

PI Young PE;

XX

XX WPI; 1999-190160/16.

DR P-PSDB; Y07763.

XX

XX New isolated human genes and the secreted polypeptides they encode

PT - useful for diagnosis and treatment of e.g. cancers, neurological

PT disorders, immune diseases, inflammation or blood disorders

XX

PS Claim 1a; Page 202; 280pp; English.

XX

XX This invention describes novel isolated human secreted proteins and

CC their encoding nucleic acid sequences. The products of the invention

CC are useful for preventing, treating or ameliorating medical conditions

CC e.g. by protein or gene therapy. Also pathological conditions can be

CC diagnosed by determining the presence or amount of expression of

CC the new polypeptides in a sample or by determining the presence or

CC absence of mutations in the new polynucleotides. Specific uses are

CC described for each of the 70 polynucleotides, based on which tissues they

CC are most highly expressed in, and include developing products for the

CC diagnosis or treatment of cancer, tumours, neurodegenerative

CC disorders, developmental abnormalities and foetal deficiencies, blood

CC disorders, leukemias, diseases of the immune system, autoimmune diseases,

CC hepatic and renal disease, lymphomas, inflammation, allergies, asthma,

CC sepsis, diabetes, Alzheimer's and cognitive disorders, schizophrenia,

THIS PAGE BLANK (USPTO)

| | | | | | |
|---------------------------|---|------------------|--------|------------|---------|
| Query Match | 8.2% | Score 124.8 | DB 2: | Length 853 | |
| Best Local Similarity | 58.2% | Pred. No. 7.3e18 | | | |
| Matches 241; Conservative | 0; | Mismatches 167; | Indels | 6; | Gaps 1. |
| 185 | gacctggccttcgtgcgcctccgcgcctcggggcgccatcgtctctgcagcgtgtcgtggg | 244 | | | |

| | | | |
|----|-----|--|-----|
| Db | 55 | GACCGGAGATTCTCTCCGACCCCTCCCGGCTTCCATCTGTCGGCGGAGATCGTTCTGGGG | 114 |
| Oy | 245 | ctctgtgtgtgtgtgtgcgtctgtattgtcgagacaacccgttaccacactgtatccgactatgtctgtg | 304 |
| Db | 115 | CTGTGGTGATATGAGCGCTTATTTGGCGAACTGAGTACTTCCGGGCCCCGGCTTTTGGCGTGG | 174 |
| Oy | 305 | gttagatgttcgcgcgtctctctcctctgtgtgtgacaatcgtccctctcaacactactgtg | 364 |
| Db | 175 | GTCATGTTGTAGCGTGAATTTTACTGGGTCCCTCAACCGCTTCTTCCTCATTTATCTACATA | 234 |
| Oy | 365 | tttcagctgcacatgaagtgtgtacatggtctccctcgtgcacatggtgtgttaatgatcttaac | 424 |
| Db | 235 | ACAATGACCTTCAACACAGGATTTCCCAAGTCCCGCGAACAACGTGGCCCTGTGCTTTAAC | 294 |
| Oy | 425 | atcagcgccacacgtctctctacataaccgcgccttaatgcgcgtctctgtggcgagtgtactgt | 484 |
| Db | 295 | GGCAATGECTTCGCTCTTGTACTCTCTGCG-----CGCTGTTGTAGATGCAATCTTCCGTC | 348 |
| Oy | 485 | acatccctcgttaggggaccccggtctataaccagcgcgcgctgtcgtctctgtcttcttgctgt | 544 |
| Db | 349 | TCCCGTGGAGGGAGACATCACAACTTCAACAGTGTGGGGGCGTCATCGTTTCTTGGCCTTC | 408 |
| Oy | 545 | tttgtgtatgatcgcgtctatagtagttagtgcctctcttcagctaccagcgcgtcgga | 598 |
| Db | 409 | CTGGTCAACATCTGCTACGCTGGAATAATCATATTTCTAGTTTATAGATGGAGCA | 462 |

```

RESULT 2
US-08-545-196B-10
; Sequence 10, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MORNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSER: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-545-196B-10

```

| | | | | |
|-----------------------|--------|--------------------|-------|--------------|
| Query Match | 4.8%; | Score 73.2; | DB 3; | Length 1582; |
| Best Local Similarity | 86.2%; | Pred. NO. 6.5e-07; | | |

| Matches | M81; Conservative | 0; Mismatches | 13; Indels | 0; Gaps |
|---------|--|---------------|------------|---------|
| Qy 1423 | actctttatttttttttaactaaataaagagcgagagaaattcaaaaaaaaaaaaaaaaa | | | 1482 |
| Db 1473 | AATTTTAAATTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA | | | 1532 |
| Qy 1483 | aaaaaaaaaaaaaaaaaaaaaaaaaaaaa | | | 1516 |
| Db 1533 | AAAAAAAAAAAAAAAAAAAAAAAAAAAAA | | | 1566 |

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

Query Match          4.8%; Score 73.2; DB 3; Length 1582;
Best Local Similarity 86.2%; Pred. No. 6.5e+07;
Matches 81; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1423 acgtgttatttttttataaaaaaagaagagagagaatataaaaaaaaaa 1482
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1473 AATATTTTATTTTATTTTAAAAAATTTTAAAAAAAAAAAAAAAAAAAA 1532

QY 1483 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1533 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1566

RESULT 4
PCT-US94-12883-3
; Sequence 3, Application PC/TUS9412883
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: UBIQUITOUS NUCLEAR RECEPTOR COMPOSITIONS AND

```


REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-702-344-26

Query Match 4.7%; Score 70.6; DB 1; Length 144;
Best Local Similarity 87.4%; Pred. No. 9.9e-07;
Matches 76; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 1430 tattttttaataaagatgaggaattaaaaa 1489
DB 13 TNNTTTTTAAAAAAAAAAAAAAAAAAAAA 72
QY 1490 aaaaaaaaaaaaaaaaaaaaaa 1516
DB 73 AAAAAAAAAAAAAAAAAAAAAA 99

RESULT 7
US-09-370-253-1
Sequence 1, Application US/09370253
Patent No. 6163792
GENERAL INFORMATION:
APPLICANT: Allen, Stephen M.
APPLICANT: Sakai, Hajime
APPLICANT: Thorpe, Catherine J.
TITLE OF INVENTION: Amino Acid Transporters
FILE REFERENCE: BB-1200
CURRENT APPLICATION NUMBER: US/09/370,253
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: 60/097,222
EARLIER FILING DATE: August 20, 1998
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Microsoft Office 97
SEQ ID NO 1
LENGTH: 1882
TYPE: DNA
ORGANISM: Hordeum vulgare
US-09-370-253-1

Query Match 4.6%; Score 70.2; DB 3; Length 1882;
Best Local Similarity 85.7%; Pred. No. 2.9e-06;
Matches 78; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 1426 gtttattttttaataaagatgaggaattaaaaa 1485
DB 1768 gtcttttcttccaaaaa 1827
QY 1486 aaaaaaaaaaaaaaaaaaaaaa 1516
DB 1828 aaaaaaaaaaaaaaaaaaaaaa 1858

RESULT 8
US-09-248-335-35
Sequence 35, Application US/09248335
Patent No. 6096504
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEF, DANIEL
TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
FILE REFERENCE: CL-1128-A
CURRENT APPLICATION NUMBER: US/09/248,335

CURRENT FILING DATE: 1999-02-10
EARLIER APPLICATION NUMBER: 08/924,759
EARLIER FILING DATE: 1997-September-05
NUMBER OF SEQ ID NOS: 74
SOFTWARE: Microsoft Word Version 7.0A
SEQ ID NO 35
LENGTH: 1098
TYPE: DNA
ORGANISM: maize
US-09-248-335-35

Query Match 4.6%; Score 70; DB 3; Length 1098;
Best Local Similarity 72.2%; Pred. No. 2.7e-06;
Matches 91; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 1391 ttgcctgcggtgtctatgtgtgtgtgtgtgtttttttaataaataa 1450
DB 957 ttgtcttcttcttattgtgtcttcaagaagtataatttccatcaaaaaa 1016
QY 1451 agatgaggaattaaaaa 1510
DB 1017 aaaaaaaaaaaaaaaaaaaaaa 1076
QY 1511 aaaaaa 1516
DB 1077 aaaaaa 1082

RESULT 9
US-08-664-596B-3
Sequence 3, Application US/08664596B
Patent No. 5807703
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavalie, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Evans, Cheryl
APPLICANT: Spaulding, Vikki
APPLICANT: Bowman, Michael
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/664,596B
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-664-5968-3

Query Match 4.5%; Score 68.6; DB 1; Length 84;
Best Local Similarity 88.1%; Pred. No. 2.2e-06;
Matches 74; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1432 tttttttaataaataagatgaggaatttaaaaaaaaaaaaaaaaaaaaaa 1491
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 TTTTNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60

OY 1492 aaaaaaaaaaaaaaaaaaaaaa 1515
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 AAAAAAAAAAAAAAAAAAAAAA 84

RESULT 10

US-08-738-367-3
Sequence 3, Application US/08738367
Patent No. 5827688

GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavallie, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vilki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/738,367
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-738-367-3

Query Match 4.5%; Score 68.6; DB 1; Length 84;
Best Local Similarity 88.1%; Pred. No. 2.2e-06;
Matches 74; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 1432 tttttttaataaataagatgaggaatttaaaaaaaaaaaaaaaaaaaaaa 1491
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 TTTTNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60

OY 1492 aaaaaaaaaaaaaaaaaaaaaa 1515
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 AAAAAAAAAAAAAAAAAAAAAA 84

RESULT 11
US-08-628-417-5
Sequence 5, Application US/08628417
Patent No. 5627054

GENERAL INFORMATION:
APPLICANT: GILLESPIE, DAVID
TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
TITLE OF INVENTION: POLYMERASE CHAIN REACTION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
CITY: ABERDEEN PROVING GROUND
STATE: MARYLAND
COUNTRY: USA
ZIP: 21010-5423

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/628,417
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BIFFONI, ULISSES J
REGISTRATION NUMBER: 39,908
REFERENCE/DOCKET NUMBER: DAM 398-94
TELECOMMUNICATION INFORMATION:
TELEPHONE: 410-671-1158
TELEFAX: 410-671-2534
INFORMATION FOR SEQ. ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 140 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: oligodeoxynucleotide
HYPOTHETICAL: NO
ANTI-SENSE: YES
US-08-628-417-5

Query Match 4.5%; Score 68.4; DB 1; Length 140;
Best Local Similarity 83.0%; Pred. No. 2.9e-06;
Matches 78; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 1423 acgtttattttttaataaataagatgaggaatttaaaaaaaaaaaaaaaaaaaaaa 1482
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 20 ACTTACAAATTAATTCTTAATAAAAAAAAAAAAAAAAAAAAAA 79

OY 1483 aaaaaaaaaaaaaaaaaaaaaa 1516
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 80 AAAAAAAAAAAAAAAAAAAAAA 113

RESULT 12

US-08-628-417-6
Sequence 6, Application US/08628417
Patent No. 5627054

GENERAL INFORMATION:
APPLICANT: GILLESPIE, DAVID
TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
TITLE OF INVENTION: POLYMERASE CHAIN REACTION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:

4.48; Score 67.4; DB 2; Length 1700;

```
Search completed: March 1, 2001, 02:07:01
Job time: 12792 sec
```

4.48; Score 67.4; DB 2; Length 1700;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2001, 19:43:14 ; Search time 8425.79 Seconds
(without alignments)
1731.065 Million cell updates/sec

Title: US-09-602-597-1
Perfect score: 2850
Sequence: 1 aagccctgaaggtlcaaaag.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 118133 seqs, 2558875100 residues
Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_bal:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_p11:*
7: gb_p12:*
8: gb_prl:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vl:*
29: gb_hc1:*
30: gb_hc2:*
31: gb_in1:*
32: gb_in2:*
33: em_bal:*
34: em_ba2:*
35: em_hum3:*
36: em_hum4:*
37: gb_prl:*
38: gb_hc3:*
39: gb_hc4:*
40: gb_hc5:*
41: gb_hc6:*
42: gb_hc7:*
43: em_hc1:*

44: em_hc2:*
45: em_hc3:*
46: em_hum5:*
47: gb_p13:*
48: gb_p15:*
49: gb_hc8:*
50: gb_hc9:*
51: gb_hc10:*
52: gb_hc11:*
53: gb_hc12:*
54: gb_hc13:*
55: gb_hc14:*
56: gb_in3:*
57: gb_hc15:*
58: gb_hc16:*
59: gb_hc17:*
60: em_hc4:*
61: em_hc5:*
62: em_hc6:*
63: em_hc7:*
64: em_hum6:*
65: gb_hc18:*
66: gb_hc19:*
67: gb_hc20:*
68: gb_hc21:*
69: gb_hc22:*
70: gb_hc23:*
71: gb_v11:*
72: gb_v12:*
73: gb_ba3:*
74: em_hc8:*
75: em_hc9:*
76: em_hc10:*
77: gb_pr6:*
78: gb_pr7:*
79: gb_sts1:*
80: gb_sts2:*
81: gb_pat1:*
82: gb_pat2:*
83: em_hc90:*
84: gb_hc24:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|---------------------|
| 1 | 2029 | 71.2 | 221262 | 30 AC009090 | AC009090 Homo sapi |
| 2 | 2029 | 71.2 | 237448 | 51 AC023825 | AC023825 Homo sapi |
| 3 | 1941 | 68.1 | 1990 | 37 AK025452 | AK025452 Homo sapi |
| 4 | 1900.4 | 66.7 | 2159 | 48 AX017393 | AX017393 Sequence |
| 5 | 1818.8 | 63.8 | 2011 | 48 AX017306 | AX017306 Sequence |
| 6 | 1505 | 52.8 | 2020 | 37 AK023291 | AK023291 Homo sapi |
| 7 | 198.8 | 7.0 | 221262 | 30 AC009090 | AC009090 Homo sapi |
| 8 | 191.4 | 6.7 | 33077 | 44 AC012092 | AC012092 Homo sapi |
| 9 | 92 | 3.2 | 7218 | 81 T66494 | Sequence 14 |
| 10 | 82 | 2.9 | 201341 | 69 AL391825 | AL391825 Homo sapi |
| 11 | 80.4 | 2.8 | 162093 | 53 AC027070 | AC027070 Homo sapi |
| 12 | 74.2 | 2.6 | 177816 | 41 AC017643 | AC017643 Drosophila |
| 13 | 74.2 | 2.6 | 262395 | 31 AE003802 | AE003802 Drosophila |
| 14 | 74 | 2.6 | 129891 | 29 AC007697 | AC007697 Drosophila |
| 15 | 73.4 | 2.6 | 3327 | 78 HSM802260 | AL137527 Homo sapi |
| 16 | 71.6 | 2.5 | 2773 | 78 HSM800446 | AL050146 Homo sapi |
| 17 | 71.4 | 2.5 | 3482 | 78 HSM800550 | AL050393 Homo sapi |
| 18 | 71 | 2.5 | 705 | 78 HSM800237 | AL049452 Homo sapi |
| 19 | 70 | 2.5 | 3787 | 78 HSM801412 | AL133565 Homo sapi |
| 20 | 69.6 | 2.4 | 363 | 32 AF135821 | AF135821 Mesobuthus |
| 21 | 69.6 | 2.4 | 3386 | 78 HSM800161 | AL050092 Homo sapi |


```
Qy 814 cactgcccctctgtcgaatgatactcgcgaacctcccaagccctggtgtcactctg 873
      |||
Db 169625 CACTGCCCTCTGCTGAGATGATGCGCATCTCCAGGCTGGGCTACTCTGGG 169684
Qy 874 agcagcagctcgtcgaatcagaagcagaagcacaatcacaatgccacggaagatt 933
      |||
Db 169685 AGCAGCAGCTCCGAGTCCAGCTCAGACAGCGAAGGACACATCAATGCGACCGGAAATTT 169744
Qy 934 gtccctccatctctcgaacacaacacctctcgaagccctctggtttctccgctac 993
      |||
Db 169745 GTCTCTCATCTTCCGACCAACACACCTCTCTGAGGCCCCCTAGTTCTCCGTCCTAC 169804
Qy 994 acaagagcctccctcccaaggttagatcgacgttcatgtcgtcctaagcatatgctc 1053
      |||
Db 169805 ACAGGAGCTCTCCCAAGGATGATCGACCTTCATGCTGCTATAGCATTTATGTC 169864
Qy 1054 cctcaaaaaaaacctctctgtcgtcatcctgtgtacaaatgacatlttaacnaatcc 1113
      |||
Db 169865 CCTCAAAAAAAACCTCTTCCCTGCAATCCGTGTACACATGACATTTTAAACCAATCC 169924
Qy 1114 aatctaaatgtgtcagaatccacacctgtgcccgaatcgtgtgtgttctctctctac 1173
      |||
Db 169925 AATCTAAAAATGTCCAGAAATCCACGTGTGCCCCGAATCGTTTGTGTTCTCTTCTAC 169984
Qy 1174 tccactcgaatgaccaaaccctgtcccgctgcacacttccctcactgatatgtgagag 1233
      |||
Db 169985 TCCACTCGAATGACCAAACTGTCCGCTGCCACTTTCCTCATGTATTTGGAGAGAG 170044
Qy 1234 gcaagcccgacggaagttccacaataaagcccgagaagaataggaacggctgtgcttg 1293
      |||
Db 170045 GCAAGGCCCGACCGCAAGTTCACACTAAAAATGCCAGGAGAAATAGGACCGCGCTGTG 170104
Qy 1294 ccaagaggtttgtgtttatgtctgttttctctccgaacgcaaaagaagta 1353
      |||
Db 170105 CCAAGAGGTTGGGTTTATGTCTCTCTGTTTCTTCTCCGACAGCAAAAGAGTA 170164
Qy 1354 aggcagctatgtgacaggtgtttaaactctatgtgtaaaatgagtggtgttgg 1413
      |||
Db 170165 AGGCAGCTTATTTGACAGGTGTATTTAAACATCTATTTGTAATGATGTGTGTGG 170224
Qy 1414 ttctactgcatgtgtgagcatgtcggtggaagaagaactgacccaggtaaatgaaatg 1473
      |||
Db 170225 TTCTACTGCAATTTGAGAGCATGCGGGGAGAGAACTGACCCAGAGTGTGTAATGAGAGCC 170284
Qy 1474 ctccctcgaactaacacagctcctgtgtgtgtgtgactaagtaaaagtataaaacccat 1533
      |||
Db 170285 CTTCCTGGAACCTAACAGTCTTGTGTGTGTGACTAACTAAAGATGATTAACCCCAT 170344
Qy 1534 ctgctgtgggtgtgacttcaacacacgcatgtcatgtgaaagctttcataccttggcc 1593
      |||
Db 170345 CTGCTGGGGGTGTCACTTTCACACTCGGCAATGTCATTTGMAAGCTTTCATACCTTGCC 170404
Qy 1594 attccctctcctcctcctccaaacccattatgtcagaagaaggtgtcctaacaaga 1653
      |||
Db 170405 ATTCCCTCTCTCTCTCTCTCCCAACCCCATTTATGACGAAAGGAGCTAAACAAGACG 170464
Qy 1654 ctcccatctcaaaacctttctctgtcgctgggaataatattatgttcttttgaataaa 1713
      |||
Db 170465 CTTCCATCTCAAACTTTTCTCTCTCTGCGGAAATTTATTTATGTTTCTTTTGAATAAAA 170524
Qy 1714 ggaattagtttaagattcctaattatagagaacaacagtagagcctgttactaataag 1773
      |||
Db 170525 GGATTAGTTTAAGATCTAAATTTTAGAGAAACAAACGTAAGGCTTGTACTAATAAGC 170584
Qy 1774 cagacatcagaactgcaggttagtgaatgaatgagatgactatlttcgcaagctctg 1833
      |||
Db 170585 CAGACATCAGAACTGACGAGTATGTTAATGATGACTTATTTCTGGAGACTCTGG 170644
Qy 1834 aatcctaataatgttaaatgtgtgagacacactgtcatatgtgtgacatcttatgagcc 1893
      |||
Db 170645 AATCTAATATTTGTAATGATGTGGACACACTTGTGATTTGTGACCATTTCTATTAGAGCC 170704
Qy 1894 c-tctcgtttaatgcatatataactgtgtctttaactgtgaaatcatatlttaacta 1952
```

```
Db 170705 CTCTCTGTTTAAAGCATATATATCTTGCTTGTAACTGTGGAATATTTAAACCTA 170764
Qy 1953 aagtgctgccctagtactttctctgtcgtcctgtgtccttttccccaacag 2012
      |||
Db 170765 AAGTGTGCTCTAGTGTCTTTCTTTCTGCTGCTGCTCTTTTCTTCTTCAACAG 170824
Qy 2013 caactctgagccatgagcagcacaacagtagaagtactgtccacactgtctcataag 2072
      |||
Db 170825 CAACTCTGAGGCCATGAGCAGCCAAACCTAAGAGTACTGCTCCACCTCGTCTATTAAG 170884
Qy 2073 gnaaacgggctcatalcccttgatctcgtgagaggaaggaaggaatggtgtgtgagccctg 2132
      |||
Db 170885 GGAACGGGCTCATCCCTTGATTTCTGAGAGAGGAGGAGAGATGTGTGTGAGAGCCTCG 170944
Qy 2133 agcagagatagatgacatgagctttgacaacaatcgttagctcctcgtcctttagaata 2192
      |||
Db 170945 AGCAGACAGATGACATGAGCTTTGACAAACATCTGTAGGCTCTCTGTTAGAAATAG 171004
Qy 2193 catgtacattcttataccatcccttattccctacaaatgttttacttcttgg 2252
      |||
Db 171005 CATGTACATCTTTATTCATCTCCCTTATTCCTACATCAATGTTTACTTCTTGGG 171064
Qy 2253 tgtgagactgagtagacacacacaaatgtgtgtgacactgtgtgtgcgagcagagagc 2312
      |||
Db 171065 TGTGAGACTGAGTAGACACACACAAATGTGTGTGACACGTGTGTGCGGAGAGAGC 171124
Qy 2313 agctactgactttgaaatgtgtgagagagccctgtgatactcatcagccactccttt 2372
      |||
Db 171125 AGCTACTGACTTTGTAAATGTGGGAGAGAGGCCCTGTGATCTCAACGACCCACCTTT 171184
Qy 2373 ccccttcagtagacactcgtgtgtccatgtgtgagatgtgagacttccctgaccca 2432
      |||
Db 171185 CCCCTTCAGTAGACGTACACTCTGTGTGCTCCATTTGCGAGATGCGCACTTCCCTCACCCA 171244
Qy 2433 taactgaatgtgtgtgaattcttctcctcttccgaactactcgtgtcactaattgtc 2492
      |||
Db 171245 TAACTGATGCTTGTGTAATCTTCTCCCTTTTGAAGACTACTGTGCTATATGTCTGC 171304
Qy 2493 cagtaatgtgtgcacagctccatccatgtgacaacaagaacattagtaaaactgtgagc 2552
      |||
Db 171305 CAGTATGTGGGCACTACACTCTCATCTCTCAACAAACAGACATTTAGTAAACTTTGTGTG 171364
Qy 2553 accttgcctctcgtcttgaattgttctgtgtgagtcgtgtgtgtgttttcaagaatg 2612
      |||
Db 171365 ACCTTGTCTCTGCTTCAATGTTCTGTGATGTCTGTGATGTCTGTGTATACAGCATGTAC 171424
Qy 2613 ccaaaacagcctcatalctgttaacagagcagcagacatcaaaatcatcattat 2672
      |||
Db 171425 CCAAAACAGCCTCACTATTGTACAGAGGAGGAGGAGCAAGACATCAATCATCTTAT 171484
Qy 2673 gtgcatgactcttaagagccattactgtatctcatgtgcctctgtatgtgaaagaagt 2732
      |||
Db 171485 GTGCGATGACTCTTAAAGGCCATTTACTGTATCTCATGAGCCTTGTGTGGAAGAAGT 171544
Qy 2733 tgaacagaggtttgagaggttttaaaacatcatataaataaagaactaaactgtcag 2792
      |||
Db 171545 TGACAGAGGTTGAGGGTTTAAAAACATCCATTAAACATGAAGCTAATAAAGCTGTGAG 171604
Qy 2793 agacaacaa 2801
      |||
Db 171605 AGAACAAGA 171613
```

```
RESULT 2
AC023825 237448 bp DNA HTG 03-SEP-2000
LOCUS Homo sapiens chromosome 16 clone RP11-322014, WORKING DRAFT
DEFINITION SEQUENCE: 20 unordered pieces.
ACCESSION AC023825
VERSION AC023825.5 GI:9965535
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
```


| | | | |
|----|-------|--|-------|
| Db | 74331 | CTGCGTGGGGGTGTCACTTCCACACTGGGCAATGATTTGTGAACCTTTCCATACCCCTTGGCC | 74330 |
| OY | 1594 | atccctctctctctctctctcccaacccttaatgacggaaggaactgcgtaaacaagc | 1653 |
| Db | 74391 | ATTCCCTCTCTCTCTCTCTCCACCCCATTTATTCAGAAAGGAGCTGCTTAACAAGAACG | 74450 |
| OY | 1654 | cttcgcatccaacctttctctgcgtctggaaattatttgattggttttgaantaat | 1713 |
| Db | 74451 | CTTCCATCTCAAACTTTCTCTGCTGGGAAATTAATTAATGTTTGGTAAATATA | 74510 |
| OY | 1714 | ggatttagtttaagaattctaaatttagaagaacaacgtaagcctgttctaataagc | 1773 |
| Db | 74511 | GGATTACTTTTAAGATTCTTAATTTTAGAGAAACAAGTAGCCCTGTGTTACTTAATAGC | 74570 |
| OY | 1774 | cagacatcaagaactgcaggtagtgatgttaatagatgaactattctctgcacgtccg | 1833 |
| Db | 74571 | CAGACATCAAGAACTTCAGAGTAGGTATGTTAATGAGATGACTTAATTTCTGGAGCTCCGG | 74630 |
| OY | 1834 | aatcctaattgttaaatgagtgaggaaacacttgatatttggaccattctcttaggc | 1893 |
| Db | 74631 | AATCCTAATATTTGTAATTAAGAGGGGACACACTGTGATTTGTGACCATTTCAATTAAGGC | 74690 |
| OY | 1894 | c-tccctgttaaatgcatatttacttgtctttaaactggaactctattcttaacta | 1952 |
| Db | 74691 | CTTCTCTGTTTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA | 74750 |
| OY | 1953 | aaggtgcgtccctagtaactttcttgtctgcctcgtcgtctcttctctccaacag | 2012 |
| Db | 74751 | AAGGTGTCGCCCTAGTACTTTTCTTGGCTGCTGCTGCTCTTCTTCTTCTTCTTCTTCTTCT | 74810 |
| OY | 2013 | caactctgagggccatgagcagcgaacaaactagagtaactgctccactcgtctctaag | 2072 |
| Db | 74811 | CAACTCTAGGCGCAATGACAGCAGCAAAACATGAGTACTGCTCCACCTGCTCTCAATAAG | 74870 |
| OY | 2073 | ggaacgggggtccatccctcttgattctctggaaggaaggaaggaagatggtgtgagggcccg | 2132 |
| Db | 74871 | GGAAAGGGGCTCATCCCTTGGAATTTGGAGGGGAGGAGGAGTGTGTGAGAGCCCTCG | 74930 |
| OY | 2133 | aggaaagagatagacatgactttgacaacaacatcgtatagctctcctgctttagaataag | 2192 |
| Db | 74931 | AGGACAGAGATAGACATGAGCTTGTGACAAACATCTGTAGGGCTCTCTCTTATGAAATAG | 74990 |
| OY | 2193 | catgtacatctcttatccattcccttatctcctaacaatglttctaacttctctgg | 2252 |
| Db | 74991 | CATGTACCATTTCTTATCCATTTCCCTTATTCCTGTACATCAATTTGTTTACTTCTTGGG | 75050 |
| OY | 2253 | tgtgaagactgagtgagacaacaacaaatgtgtgacactgtatgacggcggcaggaagc | 2312 |
| Db | 75051 | TGTGAGACTGTAGTGACACACACAAATATGTTTACACTGTGATGCCGGCAGCAGAGC | 75110 |
| OY | 2313 | agctactgaactttgaaacttggaaggaagggccctcgtatctcatccaagcccaactcttt | 2372 |
| Db | 75111 | AGCTACTGACTTTGAAACATGGGCNAGAGGCCCTGTGATTCATCCAGCCACTCTCTTT | 75170 |
| OY | 2373 | ccccctccagtaacagtgacaactctggtgccaactgagcagatgacgaactccctgaccca | 2432 |
| Db | 75171 | CCCCCTCCAGTAAAGATGACACTCTGTGTCCTCATTTGGCAGATGTGCACATTTCCCTGACCCA | 75230 |
| OY | 2433 | taacagatgactttggaattctctctctcttttcgaactactctgtgtctaattgttctgc | 2492 |
| Db | 75231 | TAACTGATGCTTTGGAATTTCTTCTCTCTTCCATTTCAACACTCTGTGTCAATTTGTTTGC | 75290 |
| OY | 2493 | cagtaaggcagcatagatccactcctctgacaacaagacatitagttaaaactttgagc | 2552 |
| Db | 75291 | CAGTATGGGCGCATAGCTCCATCTGTGACAAACAGACATTTAGSTAAACCTTTGTAGGC | 75350 |
| OY | 2553 | acctctgcgtctctgcgtcatgttctcgttgatagtcctgttgttatacagcagtgac | 2612 |
| Db | 75351 | ACCTTCTGCTTCTGCTTCAATTTCTTCTGATAGTCTCTGTTGTTATTAACAGCATGTAC | 75410 |
| OY | 2613 | ccaaacagcctcaaatgtttaaaggaaggaagggcagagaacatcaaatgcatcatcttcat | 2672 |
| Db | 75411 | CCAAACAGCCTCATATTGTTTAACAGAGGAGGGCCAGACATCAAAAGTCATCATCTTTAT | 75470 |

| | | | | | |
|-------------|---|--|--------|--------------|-------------|
| QY | 2673 | gtagcagatcactttaaagagccattacacgtatctatcattgccccttgatgttgaaaagaat | | 2732 | |
| Db | 75471 | GTGGCATGACTCTTAAAGAGCCATTACTGTATCTCATGGCCCTTGATGTGGAAAGAACT | | 75530 | |
| QY | 2733 | tgacagaggttcgcagggttttaaaaacctccattaacatgtaaacgtataataacctgtcag | | 2792 | |
| Db | 75531 | TGACAGAGGGGTTCGACAGGGGTTTAAAAACATCCATTAACTGAAGAGCTAATAAACCTGTGAG | | 75590 | |
| QY | 2793 | agaacaaaa | | 2801 | |
| Db | 75591 | AGAACACAGA | | 75599 | |
| RESULT 3 | | | | | |
| AK025452 | AK025452 | 1990 bp | mRNA | PRI | 29-SEP-2000 |
| LOCUS | Homo sapiens cDNA: FLJ21799 fis, clone HEP00611. | | | | |
| DEFINITION | AK025452 | | | | |
| ACCESSION | AK025452.1 GI:10437969 | | | | |
| VERSION | oligo capping; fis (full insert sequence). | | | | |
| KEYWORDS | Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP | | | | |
| SOURCE | clone:HEP00611. | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| TITLE | 1 (sites) | | | | |
| JOURNAL | Kawabata,A., Hiki,I.T., Kobatake,N., Inagaki,H., Ikema,Y., | | | | |
| REFERENCE | Okamoto,S., Okitani,I.R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., | | | | |
| AUTHORS | Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S. | | | | |
| | NEDO human cDNA sequencing project | | | | |
| | Unpublished (2000) | | | | |
| | 2 (bases 1 to 1990) | | | | |
| | Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., | | | | |
| | Shibahara,T., Tanaka,T. and Nakamura,Y. | | | | |
| | Direct Submission | | | | |
| | Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio | | | | |
| | Sugano, Institute of Medical Science, University of Tokyo, | | | | |
| | Laboratory of Genome Structure Analysis, Human Genome Center, | | | | |
| | Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan | | | | |
| | (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-3449-5286, | | | | |
| COMMENT | NEDO human cDNA sequencing project supported by Ministry of | | | | |
| | International Trade and Industry of Japan: cDNA full insert | | | | |
| | sequencing: Research Association for Biotechnology: cDNA library | | | | |
| | construction, 5'- and 3'-end one pass sequencing: Department of | | | | |
| | Virology and Human Genome Center, Institute of Medical Science, | | | | |
| | University of Tokyo (partly supported by Science and Technology | | | | |
| | Agency). | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..1990 | | | | |
| | /organism="Homo sapiens" | | | | |
| | /db_xref="taxon:9606" | | | | |
| | /cell_line="HepG2" | | | | |
| | /cell_type="hepatoma" | | | | |
| | /clone="HEP00611" | | | | |
| | /clone_lib="HEP" | | | | |
| | /note="Cloning vector pME18SFL3" | | | | |
| | 235..999 | | | | |
| CDS | /note="unamed protein product" | | | | |
| | /codon_start=1 | | | | |
| | /protein_id="BAB15136.1" | | | | |
| | /db_xref="GI:10437970" | | | | |
| | /translation="MDGGDSNLIILKKRFVEALIDERRKROEWEVYKRPDEPPEC | | | | |
| | PEEYIDPSLTERLOEQDKRQOEYEDQFKRWNRGIDDEPTNLDVSRQDELTER | | | | |
| | QRREELKEIKERNNLKVGISQENKKEVKRLVAKIETKRNFSQAKLLAGVKKR | | | | |
| | SGSSGNSVKRLPDPEDDKQNPESGSLGNTSISGPSIHCPSAAVCIPLGIGAM | | | | |
| | SSGSDSESSDSECTINATGKIVSISIFETNIFLEAP" | | | | |
| BASE COUNT | 562 a | 447 c | 459 g | 522 t | |
| ORIGIN | | | | | |
| Query Match | 68.1% | Score 1941; | DB 37; | Length 1990; | |

Best Local Similarity 99.8%; Pred. No. 0;
Matches 1952; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

```
OY 1 aagccctgaaggtcaaaagaatacaaaagaagaagctatttcttttttttttctt 60
Db 22 AAGCCCTGAAGGTCATAAAGAAATACAAAGCAAGGCTATTTCCTTTTTCCTTT 81
OY 61 cttcatcmttccttccttccttccttccttccttccttccttccttccttcct 120
Db 82 CTTTCATCTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 141
OY 121 aagaagagagctctcgcggttcggttcggttcggttcggttcggttcggttcg 180
Db 142 AAGAGGAGAGGCTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 201
OY 181 cctcccgagcgctgagatggttccttccttccttccttccttccttccttcct 240
Db 202 CTTCCCGCGCGGCTGAGTGGTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 261
OY 241 attatcaaaaagaggttcgttcgttcgttcgttcgttcgttcgttcgttcgttc 300
Db 262 ATTATCAAAAAGAGGTTTGTCTGAGGAGAACTAGATGAACGGCGCAAAAAGAGG 321
OY 301 gagaatgagagagagagagagagagagagagagagagagagagagagagagag 360
Db 322 GAAGAATGGAGAAAGTTGAAAGCTGAAAGCTGAAAGCTGAAAGCTGAAAGCT 381
OY 361 gacccctgagctctatgaaagagagagagagagagagagagagagagagagag 420
Db 382 GACCTCGATCTCTATATCAAAAGGCTACAGAGAACAGAGAGAGAGAGAGAG 441
OY 421 gaggaacagctcaaatcaaaacacatgtaagagagagagagagagagagagag 480
Db 442 GAGGAACAGCTCAAAATTCAAAACATGTAAAGAGGCTTATGATGAGATGAG 501
OY 481 ctgtatgaggttccttcgaagagagagagagagagagagagagagagagagag 540
Db 502 CTTGATGAGGTTTCGACAGGAGAACTAAATGAAAAACAAAGAGAGAGAGAG 561
OY 541 aagaacagagagagagagagagagagagagagagagagagagagagagagag 600
Db 562 AAGAAGCTCAAGAGATACAGAAATACCTCAAGAGAGGTTGGAATTTCTCAAG 621
OY 601 aagaagtgagaaagaagaactgactgtaagcctatagaacccaagaagaacttc 660
Db 622 AAGGAAGTGAAGAAAGAACTGACTGTAAGGCTTATGAAACCAAGAACAGTTC 681
OY 661 ggaagagctgttcgagagagagagagagagagagagagagagagagagagag 720
Db 682 GGAAGAGCTGTTGGCAGGAGCTGTGAAGCATTAAGAGCTGAGAGAGTGGCA 741
OY 721 aagaacagagagagagagagagagagagagagagagagagagagagagagag 780
Db 742 AGACTGAAGAACCGGACCTCTAGGCAAGTACAGAAATCAAGAGCCCTCATCT 801
OY 781 ctggaagaacacctcccttcgagtgagccctccatccactgcttcgagtgatc 840
Db 802 CTCGGAAGAACCTCCCTGAGTGGCCCTTCATCCACTGCTTCGTCAGATGAT 861
OY 841 ggcattccctccagagagagagagagagagagagagagagagagagagagag 900
Db 862 GGCATCTCCAGGCTGAGTGGCTCTACTCTGAGACAGGAGTCCGAGTCCAG 921
OY 901 agcgaagcaccatcatgcccacggaagatgtctccatcttcggaaccaaac 960
Db 922 AGCGAAGCACCATCAATATGCCACGGAAGATTTCTCTCTCATTTTCGAAAC 981
OY 961 ttccctgagagcccttccttccttccttccttccttccttccttccttccttc 1020
Db 982 TTCCCTGAGAGCCCTAGTTCTCGTCCCTACACAGGAGGCTCCCTCCCAAG 1041
OY 1021 cggacagctcatgctgctctatgagcattatgtccctcaaaaaaaccttcct 1080
Db 1041 cggacagctcatgctgctctatgagcattatgtccctcaaaaaaaccttccttc 1100
```

```
Db 1042 CGGACGTTTCATGCTGCTTATGAGGATTAATGTCCTCAAAAAAACTTCCTTGCCTGCA 1101
OY 1081 tccctgtacaacatgacatttcttaaccaatccatctaaatgtgcagagatccact 1140
Db 1102 TCCCTGTACAACTATGACATTTTAAACCAATTCATTTAAATATGTCAGATCCACT 1161
OY 1141 gtgagccgaatcgtgttccttccttccttccttccttccttccttccttccttc 1200
Db 1162 GTGAGCCGAATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1221
OY 1201 gctgcacatttccttccttccttccttccttccttccttccttccttccttc 1260
Db 1222 GCTGCACATTCCTCTACTATATTTGGAAGAGGAGGAGGAGGAGGAGGAGGAG 1281
OY 1261 aatgcacacagagagagagagagagagagagagagagagagagagagagagag 1320
Db 1282 AATGCCCCAGGAGAAATAGGACCGGCTGCTTCCAAAGGTTTGGTTTATTCCTTTC 1341
OY 1321 tgttttcttccttccttccttccttccttccttccttccttccttccttccttc 1380
Db 1342 TGTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 1401
OY 1381 aacacattcatgtataatgagatgtgttcgttccttccttccttccttccttc 1440
Db 1402 AAACATTCATTTGTAATGAATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1461
OY 1441 gagaagagagagagagagagagagagagagagagagagagagagagagagag 1500
Db 1462 GAAGAGAACTGACCAAGATGAATGAATGAGGCTTCCTTCGAACTAACACACTTC 1521
OY 1501 gtgtgtgactaagaaagagagagagagagagagagagagagagagagagagag 1560
Db 1522 GTTGTGTGACTAAGTAAAGATGATTAACCCCATCTGCGGGGTGCTACACTCGG 1581
OY 1561 catgattgtgaaacatttcacacatttcacacatttcacacatttcacacatttc 1620
Db 1582 CATGATTTGTGAACCTTCATACCTTCATACCTTCATACCTTCATACCTTCAT 1641
OY 1621 catctatgcaagagagagagagagagagagagagagagagagagagagagagag 1680
Db 1642 CATTTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1701
OY 1681 gggaaattatttattgtttgttttgaataaaggatttgaattgaattgaattga 1740
Db 1702 GGAATATTTATTTAGTTGTTTGAATAAAGATTTAGTTTAAAGATTTAAATTT 1761
OY 1741 ggaagaacaaagctgagctgttcttactaagccgaacatcgagagctgagtgatg 1800
Db 1762 GAGAAACAAACGTAAGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1821
OY 1801 ttaatgagatgactatttcgagagccctggaatccatattgtaattgagtgagac 1860
Db 1822 TTAATGAGATGACTTATTTTCGAGCTTCCTGGAATCTTAATATTTGTAATGAG 1881
OY 1861 acaattgatatgtgacatatttcatttgagggcc-tcttcgtttaaagcataatt 1919
Db 1882 ACACCTTCATATTTGACACATCTTATGAGGCTTCCTGTTTAAATGATATTA 1941
OY 1920 gtcctttaaactgtgagatcatttcttaacctaa 1954
Db 1942 GTGCTTTTAACGTGGAATCTATTTCTTAACCTTAA 1976
```

RESULT 4

AX017393

LOCUS

AX017393

2159 bp

DNA

PRI

07-SEP-2000

DEFINITION

Sequence 159 from Patent WO9947669.

ACCESSION

AX017393

VERSION

AX017393.1

GI:10042240

KEYWORDS

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 2159)
AUTHORS Schmitt, A., Specht, T., Dahl, E., Hinzmann, B., Rosenthal, A. and
Pillarsky, C.
TITLE Human nucleic acid sequences from tissue of breast tumors
JOURNAL Patient: WO 9947669-A 23-SEP-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES
source location/Qualifiers
1..2159
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 611 a 497 c 495 g 556 t
ORIGIN
Query Match 66.7%; Score 1900.4; DB 48; Length 2159;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1985; Conservative 0; Mismatches 11; Indels 8; Gaps 7;
QY 195 gtgattgtgttcttcattatgatgagggatgatacttatttcaaaagag 254
DB 133 GCACATTGGTTGTTCAATTGATGAGGGGATGATGTAACCTTATTCAAAAAGAG 192
QY 255 gtgtgtctgaagcagaactagatgaacgagcgcaaaagaggaagaagatggagaa 314
DB 193 GTTGTCTCTGAGGACAACTAGATGACGCGCAAAAGAGGACAGAAATGGGAGAA 252
QY 315 agtcgaaaaactgaagatccagaagaatgtccagaagaagttatgacctgactct 374
DB 253 AGTTCGAAAACCTGAAGATCCAGAGAAATGTCAGAGAGGTTTATGACCTCGATCTCT 312
QY 375 atatgaagcttcagagaagaagaagcagaagcagcagagatagagaagaagttcaa 434
DB 313 ATATGAAAGGCTACAGAAAGAGAGAGAGACAGACAGAGATGACAGAAAGTTCAA 372
QY 435 attcaaaacatgtgaagagcttagatgaagatgacacacttcttgatgagtttc 494
DB 373 ATTCAAAAATGATGTAAGAGGCTTAGTGAAGATGAGACCACTCTCTTATGAGGTTTC 432
QY 495 tcgacagcagaactaataagaagaacgaagagaagaagaactgaagaactgaaga 554
DB 433 TCGACACGAGAACTAATAGAAAAAGCAAGAGAGAAAGAACTGAAAGAACTGAAGAA 492
QY 555 atacagaataaacctcaagaaggttgaatttctcaagaagaagaagaagatgaa 614
DB 493 ATACAGAAATTAACCTCAAGAGGTTGGAATTTCTCAAGAGAACAAAGAGAGTGA 552
QY 615 gaaactgactgtgaagactatagaagaagaagaagaacttccagaagcagaagctgtg 674
DB 553 GAAACTACTGTGAGAGCTTATGAAACCAAGCAAGCAAGTTCTCCAGGCGAAG-TGTTGGC 611
QY 675 aggaagctgtgaagcataagaagctcagaagatggcacaagctgtgaagaagactga 724
DB 612 AGGAGCTGTGAGAGATTAAGAGCTCAGAGAGTGGGAACAGTGTGAAAGACTGA 671
QY 725 cccctgagccagatgaagaagaatacaagaagccctcatctgcaagttctcggaacac 794
DB 672 CCCGAGCCAGATGACAAAGATCAAGAGCCCTATCTGCAAGTCTCTCGAAACACCTC 721
QY 795 cctgagagccctccatccactgcccctgtctgtagatgtfatgagcaatctccag 854
DB 732 CCTGAGTGGCCCTCCATCCACTGCCCCCTGTGCTGCAATGATGCGACTCTCCAGG 791
QY 855 cctgggtgctactctggagcagcagctcagatccagatccagacagacgaagcaacat 914
DB 792 CCTGGGTGCTACTCTGGAGGACGCGACTCGAGTCCAGTCAAGACAGGAGCAACAT 851
QY 915 caatgcaacggagaagaatgtctctccatctccgaacaacacttctcgagggccc 974
DB 852 CAATGCCACCGAAGATGTCTCTCCATCTTCGAAACCAACACCTTCTCGAGGGCCC 911

QY 975 ctgattctccgtccctacacaggaagctctcccccagggatgatacgacgttcatgc 1034
DB 912 CTAGTTCTCTCCGTCCTACACAGGAGCTCTCCCAAGGGATGATGAGACCGTTACATGC 971
QY 1035 tgcctataggcatatgtctccctcaaaaaaactctctgctgcatctgtgtacaaca 1094
DB 972 TGCTATAGCATTAATGATCTCCCTCAAAAAAATCTCTTCTGTCATCTGTGTACACA 1031
QY 1095 tgcattttaaacaacatccatctaaagaatgtccagaatccacactgtgcccgaatcgt 1154
DB 1032 TGACATTTTAAACCAATCCATCTAAAAAATGTCCCAAAATCCACTGTGTGCCGAATCGT 1091
QY 1155 gtctgtctctcttctacccactgagatgaacaaactgtcccgctcacttctct 1214
DB 1092 GTTGGTCTCTCTTCTACTCCACTGAGATGACAAACCTGTGCCCTGCCATTTCTCT 1151
QY 1215 cactgatattggagaggaagcgaagccgaagctcgaactaaatgtccccagagaga 1274
DB 1152 CACTGATATTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1211
QY 1275 atagggcagcggtgctgtccaaaggttgggttttatgtcttctgttcttcttct 1334
DB 1212 ATAGGACACCGGCTGGCTTGGCAAAAGGTTTGGGATTTATGCTTCTGTTTCTCTTTC 1271
QY 1335 ccgacagcacaagaagaatgaaggaagttatggacaggtgttattaaacattcatgt 1394
DB 1272 CCGACACCAAAAGAAAGTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1331
QY 1395 aaatgaatgt 1454
DB 1332 AAATGAATGT 1391
QY 1455 caagtaatgaagaagagcccttccctggaactcaacagctccttgatgtgtgtgtgt 1514
DB 1392 CAGGTAATGAAGTGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1451
QY 1515 taaagatgaataaaccatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1574
DB 1452 TAAAGATGATTAACCCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1511
QY 1575 gcttccatcccttggccatcccttctctctctctctctctctctctctctctctctct 1634
DB 1512 GCTTCCATACCTTGGCCATTCCT 1571
QY 1635 gggactgtcacaagaagcgttccatctcaaaccttctctctctctctctctctctct 1694
DB 1572 GGGACTGTCTAACAAGAGAGGCTTCCATCTCAAACTTTCTCTCTCTCTCTCTCTCTCTCT 1631
QY 1695 tgttcttcttgaataaagatgttgaatgaatcctaatttgaagaagaagaacgct 1754
DB 1632 TGTTTGTTTTGAATAAAGATTTAGTTTAAGATTTCTAAATTTTGAAGAAACAAACGTA 1691
QY 1755 ggcctgttactaataagcagacatcagaactcagcagcagcagcagcagcagcagcag 1814
DB 1692 GGCCTGTGTACTTAATAGCAGACATCAGAACTGAGAGTGTATGTATGAGATGACT 1751
QY 1815 tattctgagcagctcctgtgaatccctaataatgttaaatgagtgaggacaacttcatatg 1874
DB 1752 TATTTCTGGAGGCTCTGAGATCTTAATTTGTAATGATGAGGAGACACACTTCCATATTC 1811
QY 1875 tgacattctatgagggcc-tctcgtttaaagcattatatttctgtgttcttactgt 1933
DB 1812 TGACATTTCTATTGAGGCTCTCTGTTTAATGCAATATTATATCTGTGTGTTTAACTGT 1871
QY 1934 ggaatctattctaaactaaaggtgtgcccctagtaatttc-ttgtctgctctgtcgtg 1992
DB 1872 GGAATCTATTCTTAACCTAAAGGTGCTGCCCTGATCTTTTGTGTGCTCTGCTGCTG 1931
QY 1993 tcttttctctt--ccaacagcaactctgagggcagatgagcagcaaaaactagaagttac 2050
DB 1932 TCTTTTCTCTTTTCAAAAGCAAGAACTGTGAGGCGCATGAGACAGCAAAACTGTGAGGTAC 199
QY 2051 tgtccacccctgtctataaagggaagcagcctccttgatctctgtgaggaaggaga 2110

|||||
Db 1992 TCCTCACCCTCTCATTAAGGAAACGGGCTCATCCCTGGATTTCTGAGAGAGGAGA 2051
OY 2111 gggagatggtggaagcctcgagagacagatagacagatcttgacaacatctgta 2170
|||||
Db 2052 GGGAGATGGTGTGGAGG-CTCAGAGCAGAGAGACATGAG-TTGGACACAACTCTGTA 2109
OY 2171 ggcctcctgctttagaataaga 2194
|||||
Db 2110 GGC-CGGCTGGCTTAGAATAAGCA 2132

RESULT 5
AX017306 2011 bp DNA PRI 07-SEP-2000
LOCUS AX017306 Sequence 62 from Patent WO9947669.
DEFINITION AX017306
ACCESSION AX017306
VERSION AX017306.1 GI:10042224
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Piliarsky,C.
TITLE Human nucleic acid sequences from tissue of breast tumors
JOURNAL Patent: WO 9947669-A 23-SEP-1999;
SCHMITT AKMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METZGEN GES FUER GENOMFORSCHUN
(DE); PILIARSKY CHRISTIAN (DE)
FEATURES
Source 1. 2011
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 573 a 467 c 448 g 523 t
ORIGIN

Query Match 63.8%; Score 1818.8; DB 48; Length 2011;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1868; Conservative 0; Mismatches 7; Indels 5; Gaps 4;

OY 195 gtagatgtgtgttcattatgatgatgagagatgatgtaacctattatcaaaaagag 254
|||||
Db 133 GCAGATGGGTGTTTCATTATGATGAGGAGGATGATGATACCTTATTATCAAAAAGAG 192
OY 255 gttgtgtcctgagcagacatgatgatgacgagcagaaagagagcagaagaatggagaa 314
|||||
Db 193 GTTGTGCTGTGAGGAGAACTAGATGAACGGCGCAAAAAGAGGCAAGAATGGAGAGA 252
OY 315 agttcgaagaactcgaagaatccgaagaatgtccagaagaggtttatgacctcgatcct 374
|||||
Db 253 AGTTGGAAGCCCTGAAGATCCAGGAAGATGTCACAGAGGAGTTTATGACCCCTCGATCTCT 312
OY 375 atataaagagctacagagacagagagacagagcagcagagctacagagaaagcttcaa 434
|||||
Db 313 ATATTAAGAGCTACAGAGAAACAGAGACAGAGGAGTACGAGAAACATTTCAA 372
OY 435 attcaaaaacatgylaaagagcttagatgaagatgagaccaaactccttgatgagtttc 494
|||||
Db 373 ATTCAAAAACATGTAGAGGCTTAGATGAGATGAGACCAACTTCTTGATGAGGTTTC 432
OY 495 tcgaacagcagagactaatagaaagcaacagagagagaaagaaactgaaagaa 554
|||||
Db 433 TCGACAGCAGGAATTAATTAAGAAACCAAGAGAGAAAGAACTGAAGAAGAACTGAAGA 492
OY 555 atacagaataaactcaagaagaggtggaattctcaagagaaacagaagaagagtaaaaa 614
|||||
Db 493 ATACGAATAAATCACTCAAGAGAGGTTGGAATTTCTCAAGAGAAACAGAGAGAGTGAANA 552
OY 615 gaaactgactgtaagcctatagaaacaaagacaaagttctcccaagcgaagctgtgac 674
|||||

Db 553 GAAACTGACTGTGAAGCCTATAGAAAACCAAGAACAAGTTCTCCAGCGGAG-TGTGGC 611
OY 675 aggaactgtgaagacataaagagctcagaagatggaacagttgaaagaactaaacggag 734
|||||
Db 612 AGGAGCTGTGAAGCATTAAGAGCTCAGAGAGTGGCAACAGTGAAGAACTGAACCGGA 671
OY 735 cccctgagccaga tgaacaagaatcaagagagccctcatctcgaagctctcgaagaacctc 794
|||||
Db 672 CCCTGAGCCAGATGAACAAGAAATCAAGAGCCCTCATCTGCAAGTCTCTCGAAGAACCTCC 731
OY 795 cctgagtgagccctccatccactccctctgctgcaagatgataatgcaatccctccag 854
|||||
Db 732 CCGAGAGTGGCCCTCATCTCAGTCCCTCTGCTCAGATGATGATGAGGATCTCCAGG 791
OY 855 cctgagtgagccctcctcgaagagagagcagagctcgaagctcgaagcagaagcagat 914
|||||
Db 792 CTTGGGTGCTTACTGTGGAGCAGGACTCCAGTTCAGCTAGACAGAGGAAGCACTAT 851
OY 915 caatgcacacggaagaagatgtctcctccatctccgaacacacactcctcgaagccc 974
|||||
Db 852 CAATGCCAACCAGGAAGATTGTCTCTCCATCTCCGAACACACCTTCTCGAGGCCCC 911
OY 975 ctagttctcgctccctacacagagagctcctcccaaggtagatcgagaccgtatgac 1034
|||||
Db 912 CYAGTTTCTCCCTCCCTACACAGAGAGCTCCCTCCCAAGGATGATCGACCGTTCAATGC 971
OY 1035 tccctatagagcttatgtccctcaaaaaaacctccttgctgtagatctgtgtaaca 1094
|||||
Db 972 TCCTATATGAGCATTAATGTCTCCCAAAAAAACTCTTGGCTGTGATCTGTGTAACA 1031
OY 1095 tacaattttaaacaatccaatctaataaagtgcagagatccaactgagccagatcgt 1154
|||||
Db 1032 TGACATTTTAAACAATCAATCAATTAATAATGTCAGAAATCACCTGTGCGCCGAAATGCT 1091
OY 1155 gttgtgtcctctctctactccatcagatgagacaaacgtgtccgctgcaacttctc 1214
|||||
Db 1092 GTTGTGTTCTCTCTCTACTCCATGAGATGACCAAACTGTCCCGTGCACATTTCT 1151
OY 1215 caatgatttggaagagagcgaagccgaagagctcactaaatgccccagagga 1274
|||||
Db 1152 CACTATATGTGGAGAGAGGAGAGGCCAGGCCAGGAGTTCCACTAAATATGCCCGAGAGA 1211
OY 1275 atagcagccgagctgtgcttgcgaagaaggttggtttaaattgcttctcttcttcttc 1334
|||||
Db 1212 ATAGGACACCGGTGCTGCCAAAGGTTTGGGTTTATGCTTCTGTGTTTCTTTTCTTTC 1271
OY 1335 ccgacagcacaagaagtaagagcagttatgagacagtgltatttaaacattctatgtc 1394
|||||
Db 1272 CCGACAGCACAAGAAGTAAGGGCAGTTATTGACAGGTGTTATTAAACATTTCTATTGT 1331
OY 1395 aatgaatgtgtgttggcttactatgcatgtggaagatgagggagagaaagaaactgacc 1454
|||||
Db 1332 AATGTAATGTGTTGTGTTCTACTGCAATGTGAGATGGGGGGAAGAACTGAC 1391
OY 1455 caggtaatgaaatgagccctcccttggaactaaacagcttctgtgtgtgtgactaag 1514
|||||
Db 1392 CAGGTAAATGAATGAGGCCCTTCCCTGGAACCTAACCACTGTTGATGTTGTGTGACTAAG 1451
OY 1515 taaagatgataaaccccatctgctgtgggggtgtcacttcaactcggcatgtatgtgaa 1574
|||||
Db 1452 TAAAGATGATAAACCCCATCTGCTGGGGGTGACACTTCACTCGCATGCAATTTGTGAAA 1511
OY 1575 gtttccatacccttggcattccctctctctctctctccaaacccattatagagaa 1634
|||||
Db 1512 GCTTTCATATCCCTTGGCCATTTCCCTCTCTCTCTCCAAACCCATTTATGAGAGAA 1571
OY 1635 ggaactgtcaaaagaacgcttccatccaaacacttctcctgtgagtgaaatttcta 1694
|||||
Db 1572 GGGACTGCTAACAGAAAGCGTTCCATCTCAAACTTTTCTGCTGGGAAATTAATTTTA 1631
OY 1695 tgttgtttttgaaataaaggaattagtttaagattctaaattttagaagaacaaagcga 1754
|||||
Db 1632 TGTGTTGTTTGAATTAAGGATTTAGTTTAAGATTCTAAATTTTAGAGAAACAAACGTA 1691


```
CC * 31036 31111: gap of unknown length
CC * 31112 31878: contig of 767 bp in length
CC * 31879 31954: gap of unknown length
CC * 31955 32718: contig of 764 bp in length
CC * 32719 32794: gap of unknown length
CC * 32795 32861: contig of 167 bp in length
CC * 32962 33037: gap of unknown length
CC * 33038 33077: contig of 40 bp in length.
XX
FH Key Location/Qualifiers
FH source 1. 33077
FH /db_xref="taxon:9606"
FH /organism="Homo sapiens"
FH /clone="B118"
XX
SO Sequence 33077 BP; 6694 A; 8214 C; 8318 G; 7030 T; 2821 other;

Query Match 6.7%; Score 191.4; DB 44; Length 33077;
Best Local Similarity 97.0%; Pred. No. 9.8e-29;
Matches 195; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1355 gggcagttatggacaggtgttatttaaacattctatgttaagtgtgttggc 1414
    |||
DB 31956 GCGCTGTTATGACACGCTGTTATTAACATTTGTAATGTAATGTTTGTGCT 32015

OY 1415 tctatgcatgttgagcagctggggggaagaagaaccagcgaatgaatggaccc 1474
    |||
DB 32016 TCTACTGCACTTTGTGAGCATGCGGGGGAAGAGAACTGACCCAGTAATGAAATGGAGCCC 32075

OY 1475 ttccctggaactaacacagtccttgatgtgtgtgactaagtaagaatgaataaccacatc 1534
    |||
DB 32076 TTCCCTGGAACTAACACAGCTTGATGTGTGACTAGTAAGATGATAACCCCATC 32135

OY 1535 tgcctgggggtgtcacttaaca 1555
    |||
DB 32136 TGCTGGGGGTGTCACCTCAAA 32156

RESULT 9
166494/c 166494 7218 bp DNA PAT 28-DEC-1997
LOCUS Definition Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Donner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus.
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
source 1. 7218
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 3.2%; Score 92; DB 81; Length 7218;
Best Local Similarity 6.1%; Pred. No. 1.3e-08;
Matches 26; Conservative 254; Mismatches 144; Indels 0; Gaps 0;

OY 245 tcaaaagaagttgtgtctgagcagacatagatgaacgagcgaagaagagcaagaag 304
    |||
DB 1459 TTAAGAGATGAGAAATTGCTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1400

OY 305 aatggagaaggtcgaagaacctgaagatccagaagaatgtccagaagaaggtttgacc 364
    |||
DB 1399 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1340
```

```
OY 365 ctcatctctatataagaagctacaggaacagagaagcagcagagtagcagag 424
    |||
DB 1339 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1280

OY 425 aacagttcaattcaaaaacatgtaagagcgttgatgaaatgaagaccattccttg 484
    |||
DB 1279 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1220

OY 485 atgaggttctcgacagcagcgaactaatagaaaacagcagaagaagaactgaaag 544
    |||
DB 1219 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1160

OY 545 aactgaagaatatacagaataacctaagaaggttggaaatttctcagaagaacagaag 604
    |||
DB 1159 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1100

OY 605 aagtggaaaagaactgtgactgtgaagcctatgaacccaagaacaagttcccgagga 664
    |||
DB 1099 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1040

OY 665 agct 668
    |||
DB 1039 ACCT 1036

RESULT 10
AL391825/c 201341 bp DNA HTG 06-SEP-2000
LOCUS Homo sapiens chromosome 1 clone RP11-456P18, *** SEQUENCING IN
DEFINITION PROGRAMS ***, 29 unordered pieces.
ACCESSION AL391825
VERSION AL391825.3 GI:10040176
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 201341)
AUTHORS Plumb, B.
JOURNAL Direct Submission
Submitted (06-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerquest@sanger.ac.uk
On Sep 9, 2000 this sequence version replaced gi:994368.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
Project Information
Center project name: BA456P18
----- Summary Statistics
Assembly program: XGAP; version 4.5
Sequencing vector: Plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 186960 bases at least Q40
Consensus quality: 192367 bases at least Q30
Consensus quality: 195566 bases at least Q20
Insert size: 198541; sum-of-contigs
Insert size: 216452; 5.7% error; agarose-fp
Quality coverage: 3.51x in Q20 bases; sum-of-contigs Quality
coverage: 3.36x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2205: contig of 2205 bp in length
* 2206 2305: gap of 100 bp
```

* 2306 6052: contig of 3747 bp in length
* 6053 6152: gap of 100 bp
* 6153 8931: contig of 2779 bp in length
* 8932 9031: gap of 100 bp
* 9032 15094: contig of 6063 bp in length
* 15095 15194: gap of 100 bp
* 15195 26633: contig of 11439 bp in length
* 26634 26733: gap of 100 bp
* 26734 32874: contig of 6141 bp in length
* 32875 32974: gap of 100 bp
* 32975 39594: contig of 6620 bp in length
* 39595 39694: gap of 100 bp
* 39695 42232: contig of 2538 bp in length
* 42233 42332: gap of 100 bp
* 42333 48607: contig of 6275 bp in length
* 48608 48707: gap of 100 bp
* 48708 52628: contig of 3921 bp in length
* 52629 52728: gap of 100 bp
* 52729 67221: contig of 14493 bp in length
* 67222 67321: gap of 100 bp
* 67322 73811: contig of 6490 bp in length
* 73812 73911: gap of 100 bp
* 73912 85339: contig of 11428 bp in length
* 85340 85439: gap of 100 bp
* 85440 95820: contig of 10381 bp in length
* 95821 95920: gap of 100 bp
* 95921 101396: contig of 5476 bp in length
* 101397 101496: gap of 100 bp
* 101497 108427: contig of 6931 bp in length
* 108428 108527: gap of 100 bp
* 108528 112215: contig of 3688 bp in length
* 112216 112315: gap of 100 bp
* 112316 114657: contig of 2342 bp in length
* 114658 114757: gap of 100 bp
* 114758 132207: contig of 17450 bp in length
* 132208 132307: gap of 100 bp
* 132308 148296: contig of 15989 bp in length
* 148297 148396: gap of 100 bp
* 148397 153491: contig of 5095 bp in length
* 153492 153591: gap of 100 bp
* 153592 157784: contig of 4193 bp in length
* 157785 157884: gap of 100 bp
* 157885 160643: contig of 2759 bp in length
* 160644 160743: gap of 100 bp
* 160744 165095: contig of 4352 bp in length
* 165096 165195: gap of 100 bp
* 165196 171176: contig of 5981 bp in length
* 171177 171276: gap of 100 bp
* 171277 179900: contig of 8624 bp in length
* 179901 180000: gap of 100 bp
* 180001 184362: contig of 4362 bp in length
* 184363 184462: gap of 100 bp
* 184463 191214: contig of 6752 bp in length
* 191215 191314: gap of 100 bp
* 191315 201341: contig of 10027 bp in length.

FEATURES
Source
Location/Qualifiers
1..201341
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-456P18"
/clone_1fb="RPc1-11.2"
1..2205
/note="assembly-fragment:01082
clone_end:SP6
vector_side:left"
2306..6052
/note="assembly-fragment:00659
fragment_chain:1"
6153..8931
/note="assembly-fragment:00506
fragment_chain:1"
9032..15094
misc_feature

/note="assembly-fragment:00616
fragment_chain:1"
15195..26633
/note="assembly-fragment:01635
fragment_chain:1"
26734..32874
/note="assembly-fragment:01125
fragment_chain:1"
32975..39594
/note="assembly-fragment:00639
fragment_chain:1"
39695..42232
/note="assembly-fragment:00878
fragment_chain:2"
42333..48607
/note="assembly-fragment:01344
fragment_chain:2"
48708..52628
/note="assembly-fragment:00284
fragment_chain:2"
52729..67221
/note="assembly-fragment:00228
fragment_chain:2"
67322..73811
/note="assembly-fragment:00235
fragment_chain:3"
73912..85339
/note="assembly-fragment:00329
fragment_chain:3"
85440..95820
/note="assembly-fragment:00393
fragment_chain:3"
95921..101396
/note="assembly-fragment:00629
fragment_chain:4"
101497..108427
/note="assembly-fragment:01217
fragment_chain:4"
108528..112215
/note="assembly-fragment:00389
fragment_chain:4"
112316..114657
/note="assembly-fragment:01181
fragment_chain:5"
114758..132207
/note="assembly-fragment:01139
fragment_chain:5"
132308..148296
/note="assembly-fragment:00491
fragment_chain:5"
148397..153491
/note="assembly-fragment:00556
fragment_chain:6"
153592..157784
/note="assembly-fragment:00922
fragment_chain:6"
157885..160643
/note="assembly-fragment:00452"
160744..165095
/note="assembly-fragment:00582"
165196..171176
/note="assembly-fragment:01005"
171277..179900
/note="assembly-fragment:01295"
180001..184362
/note="assembly-fragment:01314"
184463..191214
/note="assembly-fragment:01341
fragment_chain:7"
191315..201341
/note="assembly-fragment:00355
fragment_chain:7"
clone_end:T7
misc_feature


```
*          91948 92047: gap of 100 bp
*          92048 114234: contig of 22187 bp in length
*          114235 114334: gap of 100 bp
*          114335 137242: contig of 22908 bp in length
*          137243 137342: gap of 100 bp
*          137343 162093: contig of 24751 bp in length.
FEATURES
source
1. 162093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1"
/clone="RP11-195G14"
/clone_lib="NPC1-11 Human Male BAC"
1. 2191
/feature="assembly_fragment"
2292. 3768
/feature="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
3869. 6178
/feature="assembly_fragment"
6279. 11566
/feature="assembly_fragment"
11667. 16603
/feature="assembly_fragment"
16704. 21450
/feature="assembly_fragment"
21551. 28322
/feature="assembly_fragment"
28423. 35617
/feature="assembly_fragment"
35718. 47209
/feature="assembly_fragment"
47310. 57529
/feature="assembly_fragment"
57630. 73525
/feature="assembly_fragment"
73626. 91947
/feature="assembly_fragment"
clone_end:T7
vector_side:right"
misc_feature
92048. 114234
/feature="assembly_fragment"
114335. 137242
/feature="assembly_fragment"
137343. 162093
/feature="assembly_fragment"
BASE COUNT 49114 a 31642 c 30768 g 49157 t 1412 others
ORIGIN
```

```
|||||
Db 100058 AAAAAAATGTTTAA----- 10004
QY 498 acagcaggaactatagaagaagaagaagaactgaagaactgaagaata 557
Db 100042 ----AGGCTTAACCTGAAAGAAAGCATCAAGCAAGAAAGATGAGAAAGT 99988
QY 558 cagaataaactcaagaaggttggaatttcagaagaacaagaagaatggaagaag 617
Db 99987 TAGGAAAGTATCTTAACAAATGCAAGATATCT-TTGAACCAACAAAGAAATATGAGAGAA 99929
QY 618 actgactgtgaagcctaataagaacaagaagaactctccagcggaagctgtgagag 677
Db 99928 ACTGGCTGTAATTTTAAACCAAGAACTATTTCTTTAGACAAA--ATTGGCATG 99871
QY 678 agctgtgaagcataagaagctcagaaggtggaagctgtgaagaagactgaaccggacc 737
Db 99870 AACAGTCAATATTTAAAGCCCAAGAAACACAGATGATTTG-AAAGACTGAATTTGTCC 99812
QY 738 tgaagcagatgaagaagaatcaagaagccctcatctcgaagctcctcgga 787
Db 99811 TGAATCAATGTTGGAAATGAGATGTTTCATATTTGTTATTCAAA 99762
RESULT 12
LOCUS AC017643 177816 bp DNA HTG 10-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017643
VERSION AC017643.1 GI:6554354
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 177816)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
This sequence was identified as CDM:10211585 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. 177816
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 50527 a 38300 c 39085 g 49904 t
ORIGIN
```

Db 49774 GAACCCCTCAACAAATAGAGCAAGACATGAGTTGAGAGGCCCAAGCTG 49833
Oy 439 aaaca 445
| | |
Db 49834 AGTAGA 49840

RESULT 13
AE003802
LOCUS
DEFINITION Drosophila melanogaster genomic scaffold 14200001386047 section 41
of 52, complete sequence.
ACCESSION AE003802 AE002787
VERSION AE003802.2 GI:10727480
KEYWORDS
SOURCE fruit fly,
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 262395)
AUTHORS
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Stuton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D.,
Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abri,J.F., Agbayani,A., An,H.J.,
Andrews,G.L., Ahril,J.F., Baldwin,D., Balley,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Bouchan,M.R., Bouck,J., Brokstein,P., Brotlier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieres,S.,
Fleischmann,W., Foster,C., Gabriellian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennisson,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Matel,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacleb,J.M., Palazozo,M., Pittman,G.S., Pan,S.,
Pollard,D., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weissenbach,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
2 (bases 1 to 262395)
JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE

On Oct 9, 2000 this sequence version replaced gi:7302706.
Location/Qualifiers
1..262395
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

mRNA

/chromosome="2R"
<1..27
/gene="CG4878"
/product="CT43034"
/db_xref="FLYBASE:FBan0004878"
/db_xref="FLYBASE:FBgn0034237"
<1..27
/gene="CG4878"
/product="CT15671"
/db_xref="FLYBASE:FBan0004878"
/db_xref="FLYBASE:FBgn0034237"
<1..>27
/gene="CG4878"
/map="54C10-54C12"
/db_xref="FLYBASE:FBan0004878"
/db_xref="FLYBASE:FBgn0034237"
join(<246..1745,1804..>2757)
/gene="CG18432"
/product="CT41966"
/db_xref="FLYBASE:FBan018432"
/db_xref="FLYBASE:FBgn0034238"
/evidence=not_experimental
<246..>2757
/gene="CG18432"
/map="54C10-54C12"
/db_xref="FLYBASE:FBan018432"
/db_xref="FLYBASE:FBgn0034238"
/evidence=not_experimental
join(246..1745,1804..2757)
/gene="CG18432"
/note="CG18432 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan018432"
/db_xref="FLYBASE:FBgn0034238"
/evidence=not_experimental
/protein_id="AA57841.1"
/db_xref="GI:7302762"
/translation="MEALSRKTKRKRTHTTNGYRRKREOVOKLIRATEGRISPE
HDMALSLQFFHGFEGEEOENDPAVDCEDESDSGELAEPTDPSOTFPIHDLH
DPLAWMOQKORHYVDESILLYRPSNTLHFOHLSDELSKDSQIRMLNVOOSRK
SGKIGSLIQLQSGOKPASFLLMKHILRSYCHRRFOVFFPPMPSPSLKLPSPRCK
LCKQLNFSLHPQLLEHRLAQQLLELDVYDSQOKRORTRKLELEILARQALKL
ASAQDQTVAEKRLQKLTQLRLQRYVESAAQRLQLRLKLNKLELRQORFOT
RFQGLNVAHPPELAEASYCAWKSEFVDELAEVYEHLEFLYRLRLMTDOKSRFOT
GHSKPPKPPQFDRIMASLRKREYDFDPEPEYVRLADEXTARLSIPGGDLPK
ARNEFKIFPLDGOVGOSRPRYRELDQISINECTIGVLRSLSPENITVLYKSTLT
PKSRRLQISTPVLSSKDAVQKLSFQALSSQLAGDVYLYEYRSTGMSALH
DDVORLDPALRQTLPLQSLPALPQASKELVTPRPSGRIRKSQSLPIFAEQDLPFG
LDVLDLNRRLQTLRSRQQRNLHTRQKRLFPVPALEQETSEEPPLPDGNGTVQLPEGY
WNPFDLHKKHGRKFLQLLYEVIASQSAARAKALQTPLLLLAEDSRSSGSGTWTLM
RPFCSVROGGRNSLPREPSCWGHQSPDLFSCVSLHYVRAATGVPVRSHLILNLE
RRSAGGDSMSLEPVTOSE"

mRNA

join(<2926..3204,3268..3993,4054..4292,4416..>4506)
/gene="CG18631"
/product="CT41968"
/db_xref="FLYBASE:FBan018631"
/db_xref="FLYBASE:FBgn0034239"
/evidence=not_experimental
<2926..>4506
/gene="CG18631"
/map="54C10-54C12"
/db_xref="FLYBASE:FBan018631"
/db_xref="FLYBASE:FBgn0034239"
/evidence=not_experimental
join(2926..3204,3268..3993,4054..4292,4416..4506)
/gene="CG18631"
/note="CG18631 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan018631"
/db_xref="FLYBASE:FBgn0034239"
/evidence=not_experimental
/protein_id="AA57840.1"
/db_xref="GI:7302762"

CDS

1..262395
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

```

/translation="MYSNRPVLTSLSGORLCRSRTAEGSNFTWNEQLOJIOQOFGD
LRDDKISLSEDELEBOQYSDSEADYQORQCMKLEFRVPINSLASRFEQCIETAM
PRVINGYKRPILDSYTNMPTDOYPERKEVWMLXSTIEPBGDLAPLHSCALACAEK
PELOQYLDERRLELOOLPQPORYVEPLVCTAOGKRVCLTRIMDPPLPAPVASEN
PELICIRVSLSHRSYDPCGFRGVMWDNOSLDSTWCSVKDLGVLICNMLSLGL
EWMILGVSCPHGECSEVLRQPEPAELVSPAGKROLODVHCPLRITICVWGKO
NIFYNIOTETRSVMHFNLDGACMPELERRRVPAPQSCVDFDLDAKRSYSLSOLK
HIERIKKISAMRTTRKTIWNRGRYDEIFSEQLEREPNRSLSAIRFMK"
Join(6206..6413,6478..6861,6918..7090,7164..7308)
/gene="cyp33"
/Note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/Product="CT15697"
/db_xref="FLYBASE:FBgn0028382"
/db_xref="FLYBASE:FBgn0028382"
<206..7308
/gene="cyp33"
/Note="CG4886"
/map="54C11-54D1"
/db_xref="FLYBASE:FBan0004886"
/db_xref="FLYBASE:FBgn0028382"
Join(6213..6413,6478..6861,6918..7090,7164..7308)
/gene="cyp33"
/Note="cyp33 gene product; Nucleotide sequence of the
Celera sequence differs from the published sequence for
this transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0004886"
/db_xref="FLYBASE:FBgn0028382"
/protein_id="AAF57839.1"
/db_xref="GI:7302761"
/translation="MSNDRKFIYVGLADEVTERILNNAIFPGDIADIONPADYESO
RHRGAFITYESSEDAAMAIIDMNDSELCGRITRVNLAPVYKEDSEFPITADDML
OHAGATLQPESEPEAEKVEPTSPAVIAEKAKRPQVFDIRIGDNAGITVMLR
ADVPRTAENFQOLCTHEQGYKCGSFARVPEPFCOGDEFTNNNGTGKGIYKKE
NDENFLKHNSEFTLSMANGANTNGSQFICTTCTDMDNKHVHGIVGSAEYVRK
MERGKSGTSPQKIVYSCGELK"
complement(Join(7324..7910,8333..8405))
/gene="RpL18A"
/Note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/Product="CY20271"
/db_xref="FLYBASE:FBan0006510"
/db_xref="FLYBASE:FBgn0010409"
complement(<7324..7910>8405)
/gene="RpL18A"
/Note="CG6510"
/map="54C11-54D1"
/db_xref="FLYBASE:FBan0006510"
/db_xref="FLYBASE:FBgn0010409"
complement(Join(7395..7910,8333..8350))
/gene="RpL18A"
/Note="RpL18A gene product; Nucleotide sequence of the
Celera sequence differs from the published sequence for
this transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0006510"
/db_xref="FLYBASE:FBgn0010409"
/protein_id="AAF57838.1"
/db_xref="GI:7302760"
/translation="MRAGKLKEVEVGRKLPSEKPEPTLYKMRLEFADNITVARSF
WYFRLQKKFKETGELIVSIKOVYERSPVKINFGIMLEWDSRGNHMYRRYRDLTV
GCAVQCYRDMKARRHRAHNSIQIIVKDSIPAKTRRVAVKQPHDSKIFPLVQRYHH
KGRKLFSEFRKPRTFQ"

Query Match      2.6%; Score 74.2; DB 31; Length 262395;
Best Local Similarity 56.3%; Pred. No. 7e-05;
Matches 139; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

OY 199 atgtgttttcattatgatgagggatgatgtaacctattcaaaagagttt 258
      |||||
DB 25787 AATAGTTGCTGTGAAAAAGACAAATTTTACGAAAGTACGATGAGTTGCGGCTTT 25846

```

```

OY 259 gtgttcgaagcagaactagatgaacgycgcaaaagagcagaagaatggagaagt 318
      ||| |||||
DB 25847 GTACTGTAAGCTAAGCGCAGCCGACCAAGCAAGACAGACTAGAGAAATGGAGCTGTA 25906

OY 319 cgaaaactgaagaatccagaagaatgtccagaagaagttaatgacctgcattctat 378
      ||| |||||
DB 25907 CGCACCGCGAGAGATGCACATAGAACACCTGAGAACCCATGATGGCGCTCATCTTAT 25966

OY 379 gaagagctaacgaagaacagaagacgaagcagcagagatgaagaagaagttcaattc 438
      ||| |||||
DB 25967 GAACGCCCTCAACCAATAATAGCAAGAACAGACATGAGTTGAGGAAAGCCACCAAGCTG 26026

OY 439 aaaaaca 445
DB 26027 ACTAAGA 26033

RESULT 14
AC007697/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhof,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Gallego,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karia,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequerra,A., Sechi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 129891)
Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhof,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Gallego,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karia,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequerra,A., Sechi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (02-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Aug 2, 1999, this sequence version replaced gi:5549300.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
NOTE: This is a 'working draft' sequence. It currently
consists of 90 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 962: contig of 962 bp in length
* 963 1042: gap of unknown length
* 1043 2108: contig of 1066 bp in length

```

```
* 2109 2188: gap of unknown length
* 2189 2885: contig of 697 bp in length
* 2886 2965: gap of unknown length
* 2966 4067: contig of 1102 bp in length
* 4068 4147: gap of unknown length
* 4148 4828: contig of 681 bp in length
* 4829 4908: gap of unknown length
* 4909 5593: contig of 685 bp in length
* 5594 5673: gap of unknown length
* 5674 6235: contig of 562 bp in length
* 6236 6315: gap of unknown length
* 6316 7284: contig of 969 bp in length
* 7285 7364: gap of unknown length
* 7365 8030: contig of 666 bp in length
* 8031 8110: gap of unknown length
* 8111 8980: contig of 870 bp in length
* 8981 9060: gap of unknown length
* 9061 10159: contig of 1099 bp in length
* 10160 10239: gap of unknown length
* 10240 11519: contig of 1280 bp in length
* 11520 11599: gap of unknown length
* 11600 12456: contig of 857 bp in length
* 12457 12536: gap of unknown length
* 12537 13277: contig of 741 bp in length
* 13278 13357: gap of unknown length
* 13358 14226: contig of 869 bp in length
* 14227 14306: gap of unknown length
* 14307 15161: contig of 855 bp in length
* 15162 15241: gap of unknown length
* 15242 16288: contig of 1047 bp in length
* 16289 16368: gap of unknown length
* 16369 17437: contig of 1069 bp in length
* 17438 17517: gap of unknown length
* 17518 18435: contig of 918 bp in length
* 18436 18515: gap of unknown length
* 18516 19746: contig of 1231 bp in length
* 19747 19826: gap of unknown length
* 19827 20681: contig of 855 bp in length
* 20682 20761: gap of unknown length
* 20762 21781: contig of 1020 bp in length
* 21782 21861: gap of unknown length
* 21862 22251: contig of 1390 bp in length
* 22252 23331: gap of unknown length
* 23332 24141: contig of 810 bp in length
* 24142 24221: gap of unknown length
* 24222 25604: contig of 1383 bp in length
* 25605 25684: gap of unknown length
* 25685 26897: contig of 1213 bp in length
* 26898 26977: gap of unknown length
* 26978 28192: contig of 1215 bp in length
* 28193 28272: gap of unknown length
* 28273 29096: contig of 824 bp in length
* 29097 29176: gap of unknown length
* 29177 29884: contig of 708 bp in length
* 29885 29964: gap of unknown length
* 29965 31202: contig of 1238 bp in length
* 31203 31282: gap of unknown length
* 31283 32902: contig of 1620 bp in length
* 32903 32982: gap of unknown length
* 32983 34609: contig of 1627 bp in length
* 34610 34689: gap of unknown length
* 34690 36022: contig of 1333 bp in length
* 36023 36102: gap of unknown length
* 36103 37659: contig of 1557 bp in length
* 37660 37739: gap of unknown length
* 37740 38822: contig of 1083 bp in length
* 38823 38902: gap of unknown length
* 38903 40562: contig of 1660 bp in length
* 40563 40642: gap of unknown length
* 40643 41706: contig of 1064 bp in length
* 41707 41786: gap of unknown length
* 41787 43869: contig of 2083 bp in length
* 43870 43949: gap of unknown length

* 43950 46439: contig of 2490 bp in length
* 46440 46519: gap of unknown length
* 46520 49019: contig of 2500 bp in length
* 49020 49099: gap of unknown length
* 49100 51851: contig of 2752 bp in length
* 51852 51931: gap of unknown length
* 51932 53655: contig of 1724 bp in length
* 53656 53736: gap of unknown length
* 53736 55864: contig of 2129 bp in length
* 55865 55944: gap of unknown length
* 55945 58995: contig of 3051 bp in length
* 58996 59073: gap of unknown length
* 59076 62709: contig of 3634 bp in length
* 62710 62789: gap of unknown length
* 62790 67945: contig of 5156 bp in length
* 67946 68025: gap of unknown length
* 68026 71131: contig of 3106 bp in length
* 71132 71211: gap of unknown length
* 71212 75476: contig of 4265 bp in length
* 75477 75556: gap of unknown length
* 75557 79689: contig of 4133 bp in length
* 79690 79770: gap of unknown length
* 83654 83733: gap of unknown length
* 83734 88122: contig of 4389 bp in length
* 88123 88202: gap of unknown length
* 93247 93326: contig of 5044 bp in length
* 93327 101880: gap of unknown length
* 101881 101960: contig of 8554 bp in length
* 101961 102656: contig of 696 bp in length
* 102657 102736: gap of unknown length
* 102737 103419: contig of 663 bp in length
* 103420 103500: gap of unknown length
* 103500 104124: contig of 625 bp in length
* 104125 104205: gap of unknown length
* 104205 104890: contig of 686 bp in length
* 104901 104970: gap of unknown length
* 104971 105607: contig of 637 bp in length
* 105608 105687: gap of unknown length
* 105688 106356: contig of 669 bp in length
* 106357 106436: gap of unknown length
* 106437 107123: contig of 687 bp in length
* 107124 107204: gap of unknown length
* 107204 107832: contig of 629 bp in length
* 107833 107912: gap of unknown length
* 107913 108362: contig of 450 bp in length
* 108363 108442: gap of unknown length
* 108443 109244: contig of 802 bp in length
* 109245 109324: gap of unknown length
* 109325 110138: contig of 814 bp in length
* 110139 110218: gap of unknown length
* 110219 111040: contig of 822 bp in length
* 111041 111120: gap of unknown length
* 111121 111814: contig of 694 bp in length
* 111815 111894: gap of unknown length
* 111895 112635: contig of 741 bp in length
* 112636 112715: gap of unknown length
* 112716 113419: contig of 704 bp in length
* 113420 113499: gap of unknown length
* 113500 114266: contig of 767 bp in length
* 114267 114346: gap of unknown length
* 114347 115075: contig of 729 bp in length
* 115076 115155: gap of unknown length
* 115156 115792: contig of 637 bp in length
* 115793 115872: gap of unknown length
* 115873 116391: contig of 519 bp in length
* 116392 116471: gap of unknown length
* 116472 117156: contig of 685 bp in length
* 117157 117236: gap of unknown length
* 117237 117918: contig of 682 bp in length
* 117919 117996: gap of unknown length
* 117999 118596: contig of 598 bp in length
```

[illegible]

| BASE COUNT | 1004 | a | 657 | c | 636 | g | 1030 | t |
|-----------------------|---|------|--------------|---|--------------------|----------------|-----------|--------------|
| ORIGIN | | | | | | | | |
| Query Match | | | 2.6% | | Score 73.4; | | DB 78; | Length 3327; |
| Best Local Similarity | | | 70.5%; | | Pred. No. 7.3e-05; | | | |
| Matches 98; | | | Conservative | | 0; | Mismatches 41; | Indels 0; | Gaps 0; |
| QY 2712 | cctctgagtgtgaaagaagttgcacagagtgctgcaaggtcttaaaaaacatcattacat | 2771 | | | | | | |
| | | | | | | | | |
| Db 3161 | CTTTTAGTGTGTTTATATATAAGTCACAGTCAGTCTCAGACACCAAAAAAAAAAAAAAAAA | 3220 | | | | | | |
| QY 2712 | gaagctataaacctgttcagagcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa | 2831 | | | | | | |
| | | | | | | | | |
| Db 3221 | AAA | 3280 | | | | | | |
| QY 2832 | aaaaaaaaaaaaaaaaaaaaa | 2850 | | | | | | |
| | | | | | | | | |
| Db 3281 | AAAAAAAAAAAAAAAAAAAAA | 3299 | | | | | | |

```
Search completed: March 1, 2001, 02:41:45
Job time: 25111 sec
```

polyA_signal
polyA_site

GenCore Version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2001, 22:49:15 : Search time 234.97 Seconds
(without alignments)
4556.491 Million cell updates/sec

Title: US-09-602-597-1
Perfect score: 2850
Sequence: 1 aagccctgaaggtcaaaag.....aaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq-36.*
1: /SID56/gcgdata/geneseq/NA1980.DAT.*
2: /SID56/gcgdata/geneseq/NA1981.DAT.*
3: /SID56/gcgdata/geneseq/NA1982.DAT.*
4: /SID56/gcgdata/geneseq/NA1983.DAT.*
5: /SID56/gcgdata/geneseq/NA1984.DAT.*
6: /SID56/gcgdata/geneseq/NA1985.DAT.*
7: /SID56/gcgdata/geneseq/NA1986.DAT.*
8: /SID56/gcgdata/geneseq/NA1987.DAT.*
9: /SID56/gcgdata/geneseq/NA1988.DAT.*
10: /SID56/gcgdata/geneseq/NA1989.DAT.*
11: /SID56/gcgdata/geneseq/NA1990.DAT.*
12: /SID56/gcgdata/geneseq/NA1991.DAT.*
13: /SID56/gcgdata/geneseq/NA1992.DAT.*
14: /SID56/gcgdata/geneseq/NA1993.DAT.*
15: /SID56/gcgdata/geneseq/NA1994.DAT.*
16: /SID56/gcgdata/geneseq/NA1995.DAT.*
17: /SID56/gcgdata/geneseq/NA1996.DAT.*
18: /SID56/gcgdata/geneseq/NA1997.DAT.*
19: /SID56/gcgdata/geneseq/NA1998.DAT.*
20: /SID56/gcgdata/geneseq/NA1999.DAT.*
21: /SID56/gcgdata/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 1817.2 | 63.8 | 2011 | 233669 | Human breast tumor |
| 2 | 114 | 4.0 | 117 | 723042 | Human gene signal |
| 3 | 70.8 | 2.5 | 882 | 252528 | Human secreted pro |
| 4 | 68.6 | 2.4 | 2773 | 265009 | Membrane-bound pro |
| 5 | 68.4 | 2.4 | 3508 | A16619 | Human secreted pro |
| 6 | 68.2 | 2.4 | 1034 | 252527 | Human secreted pro |
| 7 | 68.2 | 2.4 | 2846 | 21 265034 | Membrane-bound pro |
| 8 | 68 | 2.4 | 1153 | 20 V55748 | Human secreted pro |
| 9 | 68 | 2.4 | 1086 | 16 087587 | DNA encoding Leuco |
| 10 | 67.8 | 2.4 | 4186 | 20 241320 | Human normal ovary |
| 11 | 67.8 | 2.4 | 2123 | 18 T59701 | PTH-like peptide D |
| 12 | 67.8 | 2.4 | 12827 | 19 V09036 | Equine arteritis v |

| | | | | | |
|----|------|-----|-------|-----------|--------------------|
| 13 | 67.8 | 2.4 | 15528 | 19 V09039 | Equine arteritis v |
| 14 | 67.6 | 2.4 | 1307 | 17 V63176 | CDNA from clone dx |
| 15 | 67.2 | 2.4 | 1641 | 19 T15758 | Human Interleukin- |
| 16 | 67 | 2.4 | 2082 | 19 V41257 | Mouse neuronal PAS |
| 17 | 66.8 | 2.3 | 2262 | 21 252560 | Human secreted pro |
| 18 | 66.6 | 2.3 | 441 | 20 209474 | p135-NT3 construct |
| 19 | 66.6 | 2.3 | 1024 | 21 297189 | Human prostate can |
| 20 | 66.6 | 2.3 | 1052 | 21 A26358 | Human secreted pro |
| 21 | 66.6 | 2.3 | 4120 | 20 209473 | Human RNA helicase |
| 22 | 66.4 | 2.3 | 2300 | 20 200686 | Human GPC3 DNA. H |
| 23 | 66.4 | 2.3 | 3637 | 20 V63190 | CDNA from clone dh |
| 24 | 66.4 | 2.3 | 3637 | 20 X60802 | Human secreted pro |
| 25 | 66.4 | 2.3 | 4061 | 19 V44867 | Clone cr797.3 codi |
| 26 | 66 | 2.3 | 1104 | 20 Z15927 | Human gene express |
| 27 | 66 | 2.3 | 1248 | 21 A26436 | Human secreted pro |
| 28 | 66 | 2.3 | 1299 | 21 A16666 | Human secreted pro |
| 29 | 66 | 2.3 | 2400 | 11 Q05055 | Placenta-specific |
| 30 | 66 | 2.3 | 3347 | 9 N80639 | Sequence of CDNA e |
| 31 | 65.8 | 2.3 | 1210 | 21 250064 | Impatiens balsamia |
| 32 | 65.8 | 2.3 | 1361 | 21 A26399 | Human secreted pro |
| 33 | 65.8 | 2.3 | 1673 | 20 X77320 | Human secreted pro |
| 34 | 65.8 | 2.3 | 4237 | 19 V61487 | Human secreted pro |
| 35 | 65.8 | 2.3 | 4650 | 20 Z33987 | Human PRO708 nucle |
| 36 | 65.8 | 2.3 | 6010 | 12 Q11998 | Human plasminogen |
| 37 | 65.8 | 2.3 | 6020 | 11 Q06648 | Plasminogen gene f |
| 38 | 65.8 | 2.3 | 7753 | 15 Q56826 | Notwalk virus stra |
| 39 | 65.6 | 2.3 | 719 | 19 V40501 | Human secreted pro |
| 40 | 65.6 | 2.3 | 772 | 19 V58363 | Coding sequence fo |
| 41 | 65.6 | 2.3 | 790 | 21 V13117 | Inserted DNA fragm |
| 42 | 65.6 | 2.3 | 790 | 21 Z39903 | DNA fragment from |
| 43 | 65.6 | 2.3 | 986 | 20 X51735 | DNA encoding a hum |
| 44 | 65.6 | 2.3 | 2685 | 21 Z43781 | Human fetal brain |
| 45 | 65.4 | 2.3 | 1954 | 21 A23441 | CDNA encoding huma |

ALIGNMENTS

| | |
|--|---------------------------------|
| RESULT 1 | |
| ID 233669 | 233669 standard; CDNA; 2011 BP. |
| AC 233669; | |
| DT 08-DEC-1999 | (first entry) |
| DE Human breast tumour-associated EST 59. | |
| EX | |
| KW Expressed sequence tag; EST; human; breast; cancer; gene therapy; | |
| XX treatment; tumour; cytostatic; medicament; ss. | |
| OS Homo sapiens. | |
| XX | |
| PN DE19813839-A1. | |
| XX | |
| PD 23-SEP-1999. | |
| XX | |
| PF 20-MAR-1998; | 98DE-1013839. |
| XX | |
| PR 20-MAR-1998; | 98DE-1013839. |
| XX | |
| PA (META-) METAGEN GES GENOMFORSCHUNG MBH. | |
| XX | |
| PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A; | |
| XX | |
| DR WPI; 1999-528981/45. | |
| XX | |
| DR P-PSDB; Y48610. | |
| XX | |
| PT Human nucleic acid sequences and protein products from tumor breast | |
| PT tissue, useful for breast cancer therapy - | |
| XX | |
| PS Claim 1a; 136-137; 188pp; German. | |
| XX | |

CC This invention describes novel human nucleic acid sequences from tumor
CC breast tissue which have cytostatic activity. The nucleic acid sequences
CC can be used to produce and isolate full-length gene sequences. They can
CC be used to express proteins, which can be used as tools to find an
CC activity against breast cancer. The sequences can be used in sense or
CC antisense form. They are especially useful for medicaments for gene
CC therapy to treat breast cancer. 33611-24617 represents expressed
CC sequence tags described in the method of the invention.

SQ Sequence 2011 BP; 573 A; 468 C; 447 G; 523 T; 0 other;

| Query Match | 63.8% | Score 1817.2 | DB 20 | Length 2011 |
|-----------------------|---|--------------|--------------|-------------|
| Best Local Similarity | 99.3% | Pred. NO. 0 | | |
| Matches 1867 | Conservative | 0 | Mismatches 8 | Indels 5 |
| | | | | Gaps 4 |
| QY 195 | gtagaatgtgttcatcatatgatagtgagggatgatgtgtaacctatcatcaaaaagag | 254 | | |
| Db 133 | gcagagtgtgttcatcatatgatagtgagggatgatgtgtaacctatcatcaaaaagag | 192 | | |
| QY 255 | gtttgtgtctcgagcgacgaactatgatgtgaacgcgcgcaaaaagaaagcaagaatctggagaa | 314 | | |
| Db 193 | gtttgtgtctcgagcgacgaactatgatgtgaacgcgcgcaaaaagaaagcaagaatctggagaa | 252 | | |
| QY 315 | agttcgaaaacctgtgaagatcccgaaagaattgtccaaagagagttttgtgccccgtctct | 374 | | |
| Db 253 | agttcgaaaacctgtgaagatcccgaaagaattgtccaaagagagttttgtgccccgtctct | 312 | | |
| QY 375 | atatgaaagcgctacacgagacaagaagaagacaagagacagagatgcagaaacattcaa | 434 | | |
| Db 313 | atatgaaagcgctacacgagacaagaagaagacaagagacagagatgcagagaaacattcaa | 372 | | |
| QY 435 | attcaaaaaacatgtgtaaagagcttatgatgtgaagtatgagaccaactctctgtatgaagtttc | 494 | | |
| Db 373 | attcaaaaaacatgtgtaaagagcttatgatgtgaagtatgagaccaactctctgtatgaagtttc | 432 | | |
| QY 495 | tcgaaacagcagaacatcatatagaagaacaaacgaagaaagaagaacatctgaaagaatgaaga | 554 | | |
| Db 433 | tcgaaacagcagaacatcatatagaagaacaaacgaagaaagaagaacatctgaaagaatgaaga | 492 | | |
| QY 555 | atacagaataaacctccaagaaggtcttggaattctctcaagaaacataagaagaatgtgnaaa | 614 | | |
| Db 493 | atacagaataaacctccaagaaggtcttggaattctctcaagaaacataagaagaatgtgnaaa | 552 | | |
| QY 615 | gaaactgacgtgtgaagcctctatagaaaaccaagaacaagtctcccaagcggaagctgtgtgc | 674 | | |
| Db 553 | gaaactgacgtgtgaagcctctatagaaaaccaagaacaagtctcccaagcggaagctgtgtgc | 611 | | |
| QY 675 | aggaagcttttgaagcctaagaagctccgaagaagtgagcaaacggtgtgnaaagaatctgaaaccgga | 734 | | |
| Db 612 | aggaagcttttgaagcctaagaagctccgaagaagtgagcaaacggtgtgnaaagaatctgaaaccgga | 671 | | |
| QY 735 | cccttgaagcctaatagtacaagaataccaagaagccctcatccctcgaagttctctcgnaaaaccttc | 794 | | |
| Db 672 | cccttgaagcctaatagtacaagaataccaagaagccctcatccctcgaagttctctcgnaaaaccttc | 731 | | |
| QY 795 | ccctgaagtgcccccttccatccactgcgcctctgtgctgagtatgtatctcggaatctctcccaag | 854 | | |
| Db 732 | ccctgaagtgcccccttccatccactgcgcctctgtgctgagtatgtatctcggaatctctcccaag | 791 | | |
| QY 855 | ccctggtgtcctaactcttggagagcagcgaactccgaftcccaagctctgaagcagcgaaagacacat | 914 | | |
| Db 792 | ccctggtgtcctaactcttggagagcagcgaactccgaftcccaagctctgaagcagcgaaagacacat | 851 | | |
| QY 915 | caatgccacccgaaagaattgtctctccatctctccgaaccacaacactctctctgagggcccc | 974 | | |
| Db 852 | caatgccacccgaaagaattgtctctccatctctccgaaccacaacactctctctgagggcccc | 911 | | |
| QY 975 | ctagtcttccgtccctcatcacagagagctctccccaaggtatagatctgagccgtttcatgc | 1034 | | |
| Db 912 | ctagtcttccgtccctcatcacagagagctctccccaaggtatagatctgagccgtttcatgc | 971 | | |
| QY 1035 | tgacctaatgacattatgtctccctcaaaaaaacctcttgcgcgtgcatctctgtgtaaca | 1094 | | |

| | | | |
|----|------|--|------|
| Db | 972 | tgctctaatgagcatatgatccctccataaaaaaacctctcttgctgcatccctggtacaa | 1031 |
| QY | 1095 | tgacaattttaaccatcccatcctaanaaalygtgcagaatgccacgtgtgcccgaatcgt | 1154 |
| Db | 1032 | tgacattttaaccatcccatcctaanaaalygtgcagaatgccacgtgtgcccgaatcgt | 1091 |
| QY | 1155 | gtttgtgtcctctcttctctactccacgtcagatgtgacaaacctgtcccgctgcacattcc | 1214 |
| Db | 1092 | gtttgtgtcctctcttctctactccacgtcagatgtgacaaacctgtcccgctgcacattcc | 1151 |
| QY | 1215 | cactgataatvgggagggcgcaaggcccaacgaaattccactaaaaatggcccagggaga | 1274 |
| Db | 1152 | cactgataatvgggagggcgcaaggcccaacgaaattccactaaaaatggcccagggaga | 1211 |
| QY | 1275 | ataagcacccggtctgagctctgcacaaaggttttgggttttattgtctctgtttttcttc | 1334 |
| Db | 1212 | ataagcacccggtctgagctctgcacaaaggttttgggttttattgtctctgtttttcttc | 1271 |
| QY | 1335 | ccgacagacacaaagaagttaaggcgagttatttgacaggtgttatttaacatccatctgt | 1394 |
| Db | 1272 | ccgacagacacacaaagaagttaaggcgagttatttgacaggtgttatttaacatccatctgt | 1331 |
| QY | 1395 | aaatgaatgtgtgtttgtgtctactgtcattgtgaggacatgvggggggaagaaactggcc | 1454 |
| Db | 1332 | aaatgaatgtgtgtttgtgtctactgtcattgtgaggacatgvggggggaagaaactggcc | 1391 |
| QY | 1455 | caggtaaatgaatgtgagccctccctccggaactaacacagctcccttgatgtgtgtgataag | 1514 |
| Db | 1392 | caggtaaatgaatgtgagccctccctccggaactaacacagctcccttgatgtgtgtgataag | 1451 |
| QY | 1515 | taaaagatgataaaccccatctgctgtgggggtgttactctcaactcggcatgcatgtgaaa | 1574 |
| Db | 1452 | taaaagatgataaaccccatctgctgtgggggtgttactctcaactcggcatgcatgtgaaa | 1511 |
| QY | 1575 | gcttccataccctctggccattccctctccctctccctcccaaccccatctttagagaa | 1634 |
| Db | 1512 | gcttccataccctctggccattccctctccctctccctcccaaccccatctttagagaa | 1571 |
| QY | 1635 | ggagactctaaacaaagacgctctccatctcaataacctttctctgctcgtggaaaattctta | 1694 |
| Db | 1572 | ggagactctaaacaaagacgctctccatctcaataacctttctctgctcgtggaaaattctta | 1631 |
| QY | 1695 | tgtttgttttggaaaataaaggatttagtttaagatccataatttagagaaacaaacgta | 1754 |
| Db | 1632 | tgtttgttttggaaaataaaggatttagtttaagatccataatttagagaaacaaacgta | 1691 |
| QY | 1755 | ggcctgtttactaataagccagacatcagagactcgcggaagtatgtaataagaagtact | 1814 |
| Db | 1692 | ggcctgtttactaataagccagacatcagagactcgcggaagtatgtaataagaagtact | 1751 |
| QY | 1815 | tattcttgcgcagctccctcgagatccctataattgttaaatgagtgagacacacatgtgcata | 1874 |
| Db | 1752 | tattcttgcgcagctccctcgagatccctataattgttaaatgagtgagacacacatgtgcata | 1811 |
| QY | 1875 | tgacacattcatatgagagcc-tctcgttttaatgcatatataacttggcttttaactgt | 1933 |
| Db | 1812 | tgacacattcatatgagagccctctcgttttaatgcatatataacttggcttttaactgt | 1871 |
| QY | 1934 | ggaatcattttcraacctaaagggtgtcgcacctgaactttc-tttgcgctcgcgcgc | 1992 |
| Db | 1872 | ggaatcattttcraacctaaagggtgtcgcacctgaactttc-tttgcgctcgcgcgc | 1931 |
| QY | 1993 | tcttttctctt--ccaaacagcaactctgagggccaatgagcccaaaaactagaaggtac | 2050 |
| Db | 1932 | tcttttctcttccaaacagcaactctgagggccaatgagcccaaaaactagaaggtac | 1991 |
| QY | 2051 | tgctccacactcgtctcataa 2070 | |
| Db | 1992 | tgctccacactcgtctcataa 2011 | |

RESULT 2

[illegible][illegible]

Db 655 ccacattccagatgaagccttcagccctctgagtcgccgacctggtgcgagtctg 714
Oy 2685 ttaagagcgcattactactgtatccatcgccctctgagtgtgaaagaattgacagaaggtt 2744
Db 715 tgagttgaccgccccagcaggtgtggtgcgtgcataatgcgtgataaagtgtgtgtca 774
Oy 2745 gcagggtttaaacaatccattcaatgaagctaataaacctgtcagagacaacaaaaa 2804
Db 775 aa 834
Oy 2805 aa 2850
Db 835 aa 880

RESULT 4
265009 standard; cDNA; 2773 BP.
XX 265009;
AC 265009;
XX 05-APR-2000 (first entry)
DF 05-APR-2000 (first entry)
XX Membrane-bound protein PRO1277 encoding cDNA.
DE Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW Pharmaceuticall; receptor immunoadhesin; gene mapping; ss.
KV
XX Homo sapiens.
OS
XX
PM WO9963088-A2.
XX 09-DEC-1999.
PD 09-DEC-1999.
XX 02-JUN-1999; 99WO-US12252.
PF 02-JUN-1999;
XX 02-JUN-1998; 98US-0087607.
PR 02-JUN-1998; 98US-0087609.
PR 02-JUN-1998; 98US-0087759.
PR 03-JUN-1998; 98US-0087827.
PR 04-JUN-1998; 98US-0088021.
PR 04-JUN-1998; 98US-0088025.
PR 04-JUN-1998; 98US-0088028.
PR 04-JUN-1998; 98US-0088029.
PR 04-JUN-1998; 98US-0088030.
PR 04-JUN-1998; 98US-0088033.
PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.

PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091518.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.


```

PR 04-JUN-1998; 98US-0088033.
PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089601.
PR 18-JUN-1998; 98US-0089607.
PR 18-JUN-1998; 98US-0089608.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090863.
PR 26-JUN-1998; 98US-0090867.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.

PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.

XX
XX
PA (GETH ) GENENTECH INC.
XX
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
XX WPI: 2000-072883/06.
DR P-PSDB: Y66695.
XX
XX Membrane-bound proteins and related nucleotide sequences
PT
XX
PS Claim 2: Fig 158; 822pp; English.
XX
CC The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor

```

immunoadhesins, for instance, can be used as therapeutic agents to block receptor-ligand interactions. The membrane-bound proteins can also be employed for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. The PRO encoding sequences are useful as hybridization probes, in chromosome and gene mapping and in the generation of antisense RNA and DNA. PRO nucleic acid sequences will also be useful for the preparation of PRO polypeptides, especially by recombinant techniques.

Sequence 2846 BP: 768 A; 696 C; 745 G; 637 T; 0 other:

Query Match 2.4%; Score 68.2; DB 21; Length 2846;
Best Local Similarity 60.5%; Pred. No. 6.3e-05;
Matches 112; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

OY 2666 tcttattatgcatgactctttaaagcgcattacgtatcattgacctgtgatgta 2725
DB 2625 tggtagcatgaccacagtaacagctgctcttccctccctcattgtacacattt 2684
OY 2726 aagaagttgacagaggttgcaggtttaaacatccattacattgaagtaaac 2785
DB 2685 aataaataaaggttgcgtctcgaactacataaaaaaaataaaaaaaataaaaa 2744
OY 2786 ctgtcagagaaacaaaaaaataaaaaaaataaaaaaaataaaaaaa 2845
DB 2745 aaaaaaaataaaaaaaataaaaaaaataaaaaaaataaaaaaaataaaaa 2804
OY 2846 aaaa 2850
DB 2805 aaaa 2809

RESULT 8
V55748
ID V55748 standard; cDNA; 1153 BP.

AC V55748;
XX
DT 23-MAR-1999 (first entry)

DE Human secreted protein clone ej265_4 coding sequence.

KW Secreted protein; human; nutritional supplements; immune stimulant;
KW Immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;
KW activin/inhibin; chemokinesis; haemostasis; thrombolysis;
KW receptor/ligand activity; anti-inflammatory; tumour inhibitor;
KW cadherin/tumour invasion suppressor; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
XX CDS 261..899
XX FT /*tag= a

XX MO9855614-A2.

XX 10-DEC-1998.

XX 01-JUN-1998; 98WO-US11210.

XX 29-MAY-1998; 98US-0087255.
XX 04-JUN-1997; 97US-0868696.
XX 04-JUN-1997; 97US-0868697.
XX 04-JUN-1997; 97US-0868698.
XX 04-JUN-1997; 97US-0868898.
XX 04-JUN-1997; 97US-0868899.
XX 04-JUN-1997; 97US-0868900.
XX 04-JUN-1997; 97US-0869191.
XX 04-JUN-1997; 97US-0869192.
XX 04-JUN-1997; 97US-0869193.
XX 04-JUN-1997; 97US-0869194.

PA (GENY) GENETICS INST INC.

XX Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;

PI McCoy JM, Racie LA, Spaulding V, Treacy M;

XX WPI: 1999-059912/05.

DR P-PSDB: W73630.

PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
PT retina, adult placenta or adult uterus cDNA libraries

XX Claim 32; Page 92-93; 127pp; English.

CC This sequence encodes a human secreted protein of the invention.
CC This DNA sequence was isolated from a human adult placenta cDNA
CC library, and was designated clone ej265.4. The DNAs and proteins
CC are predicted to have biological activities which would make them
CC suitable for treating, preventing or ameliorating medical conditions in
CC humans and animals, although no supporting data is given. Suggested
CC activities include nutritional sources or supplements, immune
CC stimulating or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity. The DNAs
CC are also stated to be useful for gene therapy. A host cell transfected
CC with the DNA, or its subfragments and variants is useful for recombinant
CC production of the human secreted protein clones.

XX Sequence 1153 BP: 346 A; 277 C; 232 G; 298 T; 0 other:

Query Match 2.4%; Score 68; DB 20; Length 1153;
Best Local Similarity 80.0%; Pred. No. 5.2e-05;
Matches 80; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 2751 tttaaacatcattacatgaagctaaacctgtcagagaaacaaaaaaata 2810

DB 1054 ttatattcatgtgcaaacacaaaaataaataattctgcatgcataaaaaaaata 1113

OY 2811 aaaaaaaataaaaaaaataaaaaaaataaaaaaaata 2850
DB 1114 aaaaaaaataaaaaaaataaaaaaaataaaaaaaata 1153

RESULT 9
O87587
ID O87587 standard; DNA; 1686 BP.

XX O87587;

XX 19-DEC-1995 (first entry)

DE DNA encoding Leucocytozoan protozoa structural protein epitope.

XX Leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;

XX Leucocytozoanosis; treatment; ss.

XX Leucocytozoan protozoa sp.

XX JP07089995-A.

XX 04-APR-1995.

XX 10-SEP-1993; 93JP-0226078.

XX 10-SEP-1993; 93JP-0226078.

XX (DOBU-) DOBUTSUVO SEIBUTSUGAKOTEKI SRIZAI KYOKAI.
XX (NITS-) NITSEIKEN KK.
XX WPI: 1995-167252/22.

| | | | | | | |
|-----------------------|---|-----------|-------------|------------|-------------|---------|
| Pt | Rosenthal A, | Specht T, | Hinzmann B, | Schmitt A, | Pilarsky C, | Dahl E; |
| Xx | wpl; 1999-552352/47. | | | | | |
| XX | Nucleic acid sequences potentially useful in diagnosis or therapy of ovarian cancer - | | | | | |
| PS | Claim 3; Page 209-210; 274ppp; German. | | | | | |
| xx | This invention describes novel nucleic acid sequences that are highly expressed in normal ovary tissue. Artificial chromosomes and cosmid clones containing the sequences can be used as gene transfer vehicles. The sequences can be used to produce DNA fragments containing full-length genes. Host cells transformed with the sequences can be used to produce polypeptides or polypeptide fragments, which can be used to screen phage displays for polypeptides that bind to them, or as tools for identifying agents active against ovarian cancer, or to prepare medicaments for treating ovarian cancer. The cdna sequences can be used to obtain genomic genes, their promoters, enhancers, silencers, exon structures, intron structures and their splice variants. 241222-241324 represent cDNA sequences derived from normal human ovarian tissue and which encode the protein fragments represented in Y59724-Y59837. | | | | | |
| SO | Sequence 4186 BP; 938 A; 1199 C; 1234 G; 815 T; 0 other; | | | | | |
| Query Match | 2.4%; Score 68; DB 20; Length 4186; | | | | | |
| Best Local Similarity | 76.9%; Pred. No. 7.8e-05; | | | | | |
| Matches | 83; Conservative 0; Mismatches 25; Indels 0; Gaps 0; | | | | | |
| Oy | 2743 ttgcaggtttaaaacatccattgaagaacttaataaacctgtcagagacaanaa 2802 Db 2590 ttgaagtgtgtcaaaaaaaaiaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2649 | | | | | |
| Oy | 2803 aa 2850 Db 2650 aa 2697 | | | | | |
| RESULT .11 | | | | | | |
| T59701 | | | | | | |
| XX | T59701 standard; DNA: 2123 BP. | | | | | |
| XX | T59701; | | | | | |
| DT | 07-JUL-1997 (first entry) | | | | | |
| XX | PTH-like peptide DNA clone HHM-4. | | | | | |
| DE | PTH-like peptide; parathyroid hormone; parathormone; agonist; KM adenylate cyclase; humoral hypercalcaemia of malignancy; HHM; KW tumour marker; cancer; diagnosis; ss. | | | | | |
| XX | Homo sapiens. | | | | | |
| OS | | | | | | |
| FH | Key Location/Qualifiers | | | | | |
| FT | 5'UTR 1..938 /tag= a | | | | | |
| FT | CDS 939..1547 /tag= b | | | | | |
| FT | sig_peptide 939..1046 /tag= c | | | | | |
| FT | mat_peptide 1047..1544 /tag= d | | | | | |
| FT | 3'UTR 1545..2123 /tag= e | | | | | |
| FT | polyA_signal 2006..2011 /tag= f | | | | | |
| FT | misc_feature 1062..1099 /tag= g | | | | | |
| TT | /note= "nucleotide region corresponding to probe used to screen SKRC-1 CDNA library" | | | | | |
| XX | | | | | | |


```
XX 14-JUL-1997; 97WO-NL00408.
PF 12-JUL-1996; 96NL-0003579.
XX (UYLE-) RIJKSUNIV LEIDEN.
XX Bredenbeek PJ, Den Boon JA, Snijder EJ, Spaan WM;
PI Van Dinten LC, Wassenaar ALM;
XX WPI; 1998-110596/10.
DR
XX Recombinant DNA vector that produces infectious RNA of equine
PT arteritis virus - and derived RNA, host cells and virus particles,
PT useful therapeutically
XX
XX Claim 4; Pages 21-24; 36pp; English.
XX
XX The sequence is that of of a recombinant DNA vector, pEAV030,
CC comprising a DNA sequence complementary to part of
CC the genome of equine arteritis virus (EAV) which has
CC adjacent to its 3'-end a non-EAV specific promoter for a non-EAV
CC specific RNA polymerase. It can be used to make viral particles
CC and for therapeutic use. The vector can be stably incorporated
CC into host cells and produces infectious RNA.
XX
SQ Sequence 15528 BP; 3492 A; 3941 C; 3972 G; 4123 T; 0 other;

Query Match 2.4%; Score 67.8; DB 19; Length 15528;
Best Local Similarity 62.9%; Pred. No. 0.00013;
Matches 105; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 2684 ctaagaagccatctactgtatctcatgycctcttgatgtaagaagtgacagaaggt 2743
DB 12641 cgtaaagacggtgattctctctgtgctgcatgttgagtagtattagccaccag 12700
OY 2744 tgcaggggtttaaaacatcattcaatgaagctataaacctgtcagagacaacaaa 2803
DB 12701 aaccacaaaaaaacacacacacacacacacacacacacacacacacacacacac 12760
OY 2804 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2850
DB 12761 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 12807

RESULT 14
V63176
ID V63176 standard; CDNA: 1307 BP.
XX
AC V63176;
DT 12-JAN-1999 (first entry)
DE CDNA from clone dx279_1 which encodes a secreted protein.
XX
KW Secreted protein; immune stimulating; suppressing;
KW haematopoiesis regulating activity; tissue growth activity; activin;
KW inhibin activity; chemotactic; chemokinetic activity; haemostatic;
KW thrombolytic activity; anti-inflammatory activity; cadherin;
KW tumour invasion suppressor activity; tumour inhibition activity; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 449..1009
XX FT /*tag= a
XX PN WO9840486-A2.
XX PD 17-SEP-1998.
XX PF 13-MAR-1998; 98WO-US04977.
```

```
XX 29-OCT-1997; 97US-0960022.
PR 14-MAR-1997; 97US-0815047.
XX (GEMV ) GENETICS INST INC.
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racle LA, Spaulding V, Treacy M;
XX WPI; 1998-520812/44.
DR P-PSDB; W80402.
XX
XX New isolated human poly(nucleotide)s and secreted proteins -
PT obtained from e.g. human foetal kidney, placenta, foetal brain,
PT adult testes, adult brain or adult uterus CDNA libraries
XX
XX Claim 35; Pages 84-85; 110pp; English.
XX
XX The present sequence encodes a secreted protein. The nucleic acid
CC sequence is isolated from a human adult testes CDNA library using
CC probe V63186. The polypeptide may have biological activities such as
CC e.g. nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity or other activities.
XX
SQ Sequence 1307 BP; 295 A; 442 C; 370 G; 200 T; 0 other;

Query Match 2.4%; Score 67.6; DB 19; Length 1307;
Best Local Similarity 80.6%; Pred. No. 6.5e-05;
Matches 79; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 2752 ttaaaacatccattacttgaaagctaaacacctgtcagagacacacacacacacacac 2811
DB 1210 ttaaaatcacctgttgcaaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 1269
OY 2812 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2849
DB 1270 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 1307

RESULT 15
T15758
ID T15758 standard; CDNA: 1641 BP.
XX
AC T15758;
DT 24-JUL-1997 (revised)
DT 25-JUN-1996 (first entry)
DE Human Interleukin-15 receptor CDNA from clone p1.
XX
KW Interleukin-15 receptor; IL-15R; murine; human; graft rejection;
KW treatment; prevention; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 3..842
XX FT /*tag= a
XX FT /product= interleukin-15_receptor
XX PN WO9530695-A.
XX PD 16-NOV-1995.
XX PF 04-MAY-1995; 95WO-US05585.
XX PF 06-SEP-1994; 94US-0300903.
```

PR 06-MAY-1994; 94US-0236919.

XX (IMV) IMMUNEX CORP.

XX Anderson DM, Giri JG;

XX WPI; 1996-077255/08.

DR P-PSDB; R90843.

XX DNA encoding an interleukin-15 receptor - used to ameliorate undesired
PT effects of interleukin-15, e.g. to prevent or treat graft rejection

XX Example 3; Page 35-37; 53pp; English.

XX A lambda-gt10 library from human peripheral blood lymphocytes, was
CC screened for the presence of a full-length clone encoding human
CC interleukin-15 receptor (IL-15R) using a random prime labeled human
CC IL-15R cDNA probe consisting of the entire W5 cDNA clone (799999)
CC without the polyA tail. The resulting cDNA insert from this clone
CC T15758 encodes R90843, and exhibited an in-frame insertion of 153 bp
CC at the mature N-terminus (amino acids 24 through 74), an in-frame
CC deletion of 99 bp downstream of the mature N-terminus that deleted
CC nucleotides 236 to 334 of T15995, and also contained additional 5'
CC sequence as compared to clone W5 (amino acids 1 through 10), but still
CC did not contain an initiator methionine. The IL-15R DNA is a multi-exon
CC gene. IL-15R variants can be attributed to different mRNA splicing events
CC following transcription or from proteolytic cleavage of the IL-15R
CC protein, where the IL-15R binding property is retained. The IL-15R
CC protein is useful in treatment of undesired effects of interleukin-15,
CC e.g. to prevent or treat graft rejection.
CC (Revised entry submitted to correct sequence analysis breakdown.)

XX Sequence 1641 BP; 478 A; 463 C; 385 G; 315 T; 0 other;

XX Query Match 2.4%; Score 67.2; DB 17; Length 1641;

XX Best Local Similarity 81.2%; Pred. No. 8.5e-05;

XX Matches 78; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 2755 aaacatcattacatgaagctataaactctcagagaaacaaaaaaaaaaaaa 2814

Db 1535 aaacatcattcattcaataaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1594

OY 2815 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2850

Db 1595 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1630

Search completed: March 1, 2001, 02:11:11
Job time: 12116 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2001, 18:20:14 ; Search time 3991.74 Seconds
(without alignments)
5003.167 Million cell updates/sec

Title: US-09-602-597-1
Perfect score: 2850
Sequence: 1 aagccctgaaggtcaaaag.....aaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_est1a:*
42: em_est1b:*
43: em_est1c:*

44: em_est1d:*
45: em_est1e:*
46: em_est1f:*
47: em_est1g:*
48: em_est1h:*
49: em_est1i:*
50: em_est1j:*
51: em_est1k:*
52: em_est1l:*
53: em_est1m:*
54: em_est1n:*
55: em_est1o:*
56: em_est1p:*
57: em_est1q:*
58: em_est1r:*
59: em_est1s:*
60: em_est1t:*
61: em_est1u:*
62: em_est1v:*
63: em_est1w:*
64: em_est1x:*
65: em_est1y:*
66: em_est1z:*
67: em_est2a:*
68: em_est2b:*
69: em_est2c:*
70: em_est2d:*
71: em_est2e:*
72: em_est2f:*
73: em_est2g:*
74: em_est2h:*
75: em_est2i:*
76: em_est2j:*
77: em_est2k:*
78: em_est2l:*
79: em_est2m:*
80: em_est2n:*
81: em_est2o:*
82: em_est2p:*
83: em_est2q:*
84: em_est2r:*
85: em_est2s:*
86: em_est2t:*
87: em_est2u:*
88: em_est2v:*
89: em_est2w:*
90: em_est2x:*
91: em_est2y:*
92: em_est2z:*
93: em_est3a:*
94: em_est3b:*
95: em_est3c:*
96: em_est3d:*
97: em_est3e:*
98: em_est3f:*
99: em_est3g:*
100: em_est3h:*
101: em_est3i:*
102: em_est3j:*
103: em_est3k:*
104: em_est3l:*
105: em_est3m:*
106: em_est3n:*
107: em_est3o:*
108: em_est3p:*
109: em_est3q:*
110: em_est3r:*
111: em_est3s:*
112: em_est3t:*
113: em_est3u:*
114: em_est3v:*
115: em_est3w:*
116: em_est3x:*

117: em_estp16:*

118: em_estp17:*

119: em_estp18:*

120: em_estp19:*

121: em_estp20:*

122: em_estp21:*

123: em_estp22:*

124: em_estp23:*

125: em_estp24:*

126: em_estp25:*

127: em_estp26:*

128: em_estp27:*

129: em_estp28:*

130: em_estp29:*

131: em_estp30:*

132: em_estp31:*

133: em_estp32:*

134: em_estp33:*

135: em_estp34:*

136: em_estp35:*

137: em_estp36:*

138: em_estp37:*

139: em_estp38:*

140: em_estp39:*

141: em_estp40:*

142: em_estp41:*

143: em_estp42:*

144: em_estp43:*

145: em_estp44:*

146: em_estp45:*

147: em_estp46:*

148: em_estp47:*

149: em_estp48:*

150: em_estp49:*

151: em_estp50:*

152: em_estp51:*

153: em_estp52:*

154: em_estp53:*

155: em_estp54:*

156: em_estp55:*

157: em_estp56:*

158: em_estp57:*

159: em_estp58:*

160: em_estp59:*

161: em_estp60:*

162: em_estp61:*

163: em_estp62:*

164: em_estp63:*

165: em_estp64:*

166: em_estp65:*

167: em_estp66:*

168: em_estp67:*

169: em_estp68:*

170: em_estp69:*

171: em_estp70:*

172: em_estp71:*

173: em_estp72:*

174: em_estp73:*

175: em_estp74:*

176: em_estp75:*

177: em_estp76:*

178: em_estp77:*

179: em_estp78:*

180: em_estp79:*

181: em_estp80:*

182: em_estp81:*

183: em_estp82:*

184: em_estp83:*

185: em_estp84:*

186: em_estp85:*

187: em_estp86:*

188: em_estp87:*

189: em_estp88:*

190: gb_gss25:*

191: gb_gss26:*

192: gb_gss27:*

193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|-----|----------|--------------------|
| C 1 | 699.8 | 24.6 | 704 | 23 | A1655644 | A1655644 ts99g02.x |
| C 2 | 690.4 | 24.2 | 712 | 158 | A0470436 | A0470436 CTBT-F1- |
| C 3 | 606 | 21.3 | 632 | 135 | BE72768 | BE72768 601564388 |
| C 4 | 596.2 | 20.9 | 614 | 95 | AM821142 | AM821142 PM2-ST030 |
| C 5 | 594.2 | 20.9 | 638 | 107 | BE387718 | BE387718 601287377 |
| C 6 | 594.2 | 20.8 | 917 | 135 | BE732904 | BE732904 601567847 |
| C 7 | 588.2 | 20.6 | 699 | 136 | BE869679 | BE869679 601445783 |
| C 8 | 583.4 | 20.5 | 664 | 107 | BE392702 | BE392702 601307895 |
| C 9 | 581.8 | 20.4 | 631 | 105 | BE180639 | BE180639 RC3-HT062 |
| C 10 | 575.2 | 20.2 | 895 | 135 | BE736929 | BE736929 601306864 |
| C 11 | 564.8 | 19.8 | 568 | 89 | AM327476 | AM327476 dq03g09.x |
| C 12 | 552.4 | 19.4 | 554 | 92 | AM592821 | AM592821 hq06c01.x |
| C 13 | 534.4 | 18.8 | 568 | 17 | A1205174 | A1205174 ap13a11.x |
| C 14 | 533.2 | 18.7 | 577 | 15 | A1031969 | A1031969 ovs5d01.x |
| C 15 | 532.4 | 18.7 | 623 | 38 | AV653573 | AV653573 AV653573 |
| C 16 | 519 | 18.2 | 519 | 136 | BE858277 | BE858277 7918g01.x |
| C 17 | 501.8 | 17.6 | 514 | 40 | AM119015 | AM119015 xq99b10.x |
| C 18 | 499 | 17.5 | 510 | 12 | AA854150 | AA854150 aj71f04.s |
| C 19 | 499 | 17.5 | 529 | 17 | A1188440 | A1188440 qd14b03.x |
| C 20 | 492.4 | 17.3 | 495 | 38 | AM004593 | AM004593 w890a04.x |
| C 21 | 489 | 17.2 | 489 | 26 | A1863261 | A1863261 tz45c12.x |
| C 22 | 487.8 | 17.1 | 491 | 9 | AA594110 | AA594110 tmd1e09.s |
| C 23 | 482 | 16.9 | 482 | 19 | A1346704 | A1346704 qp47b12.x |
| C 24 | 481.6 | 16.9 | 627 | 95 | AM821178 | AM821178 PM2-ST030 |
| C 25 | 476 | 16.7 | 476 | 25 | A1819378 | A1819378 w61d09.x |
| C 26 | 475 | 16.7 | 479 | 16 | A1090352 | A1090352 oy18b10.s |
| C 27 | 474.4 | 16.6 | 476 | 16 | A1146389 | A1146389 q805c07.x |
| C 28 | 469 | 16.5 | 481 | 105 | BE253505 | BE253505 601108191 |
| C 29 | 467.8 | 16.4 | 471 | 25 | A1799081 | A1799081 we98d07.x |
| C 30 | 464.4 | 16.3 | 469 | 89 | AM361848 | AM361848 PM3-CF026 |
| C 31 | 464 | 16.3 | 465 | 25 | A1819590 | A1819590 wk03a12.x |
| C 32 | 460.4 | 16.2 | 462 | 13 | AA873225 | AA873225 oh70g08.s |
| C 33 | 456 | 16.0 | 457 | 40 | AM117393 | AM117393 x800f07.x |
| C 34 | 452.8 | 15.9 | 499 | 39 | AM084108 | AM084108 xc37c08.x |
| C 35 | 450 | 15.8 | 475 | 89 | AM327477 | AM327477 dq03g09.y |
| C 36 | 449 | 15.8 | 449 | 21 | A1479864 | A1479864 tm0c12.x |
| C 37 | 448.8 | 15.7 | 472 | 137 | BE393496 | BE393496 OVO-UM009 |
| C 38 | 448 | 15.7 | 448 | 134 | BE047617 | BE047617 tz39e03.y |
| C 39 | 448 | 15.7 | 471 | 142 | H99248 | H99248 yx21g06.s1 |
| C 40 | 447.4 | 15.7 | 449 | 22 | A1613225 | A1613225 cy39g01.x |
| C 41 | 443 | 15.5 | 443 | 17 | A1192097 | A1192097 q606c09.x |
| C 42 | 442 | 15.5 | 444 | 9 | AA598697 | AA598697 aed4b05.s |
| C 43 | 441 | 15.5 | 453 | 16 | A1148841 | A1148841 qc65d12.x |
| C 44 | 439 | 15.4 | 439 | 18 | A1278270 | A1278270 q178b06.x |
| C 45 | 437 | 15.3 | 437 | 19 | A1359389 | A1359389 q728b06.x |

ALIGNMENTS

RESULT 1

LOCUS A1655644

DEFINITION ts99g02.x1 NC1 CGAP GC6 Homo sapiens cDNA clone IMAGE:2239442 3'

ACCESSION A1655644

VERSION A1655644.1

KEYWORDS EST.

SOURCE human.

704 bp mRNA

GI:4739623

14-DEC-1999

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 704)
 NCBI-Genbank http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1350
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCBI-Genbank clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLNL at:
 www.bio.lml.gov/dbp/image/image.html
 Inset length: 844 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 467.
 Location/Qualifiers
 1..704
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="2239442"
 /clone_lib="NCI-CGAP-GC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Plasmid DNA from the normalized library
 NCI-CGAP-GC4 was prepared, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from a pool of 5,000 clones made
 from the same library (clones: 1257096-1258631,
 1469064-1470983, and 1475592-1476743). Subtraction by
 Bento Soares and M. Fatima Bonaldo."
 -BASE COUNT 139 a 156 c 186 g 222 t 1 others
 ORIGIN
 Query Match 24.6%; Score 699.8; DB 23; Length 704;
 Best Local Similarity 99.6%; Pred. No. 4e-141;
 Matches 701; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 344 TCCTGATGTCGCCCTTCATCCACTGCCCCCTGCTGCTCATATGATGCGCATCTCCCA 285
 QY 853 ggcctgggtgctactcttggaagcagcactccgagtcagtcagcagcaggaagcacc 912
 Db 284 GGCTTGATGCTACTCTGGAGCAGGACTCCGAGTCAGTCAGACAGGAGCACC 225
 QY 913 ataatgcccgcggaagatgtgtctctcatcttcgaaccacaccttcctcgagcc 972
 Db 224 ATCAATGCCACCGGAAGATGTCCTTCATCTCCGAACCAACCTTCGAGGCC 165
 QY 973 ccttagttctcgtccctcacagagagctctcccaagggaagtcgagccgttcat 1032
 Db 164 CCTAGTTCTCGCTCCCTACACAGGAGCTCCTCCCAAGGTAGTCGAGCCGTTTCA 103
 QY 1033 gctgcctaatagcattatgtccctcaaaaaaactccttgctgcatctgtgtacaa 1092
 Db 104 GCTGCTATAGGATATGTCCTCAAAAAAAACCTTCCTGCTGATCCTGTACAA 45
 QY 1093 catgacatttttaaccatccatccataaaatgtgcagatcc 1136
 Db 44 CATGACATTTTAAACCAATCAATCAATAAATGTCGACAGATCC 1

RESULT 2
 AQ470436/c 712 bp DNA GSS 23-APR-1999
 LOCUS CTRBI-El-2584F21.TF CTRBI-El Homo sapiens genomic clone 2584F21,
 DEFINITION DNA sequence.
 ACCESSION AQ470436
 VERSION AQ470436.1 GI:4654090
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 712)
 AUTHORS Zhao,S., Adams,M.D., Nieman,W., Malek,J., Shizuya,H., Simon,B. and
 Venter,J.C.
 TITLE Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
 Map Building
 JOURNAL Unpublished (1997)
 COMMENT other_GSSs: CTRBI-El-2584F21.TF
 Contact: Shaying Zhao, William Nieman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0208
 Fax: 301 838 0208
 Email: hbeet@igf.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
 http://www.tigr.org/tdb/hungen/Bac_end_search/bac_end_search.html.
 Seq primer: M13-21
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..712
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="2584F21"
 /clone_lib="CTRBI-El"
 /sex="male"
 /cell_type="sperm"
 /note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
 Caltech Human BAC Library D"

BASE COUNT 210 a 157 c 171 g 174 t
 ORIGIN
 Query Match 24.2%; Score 690.4; DB 158; Length 712;
 Best Local Similarity 99.0%; Pred. No. 4.3e-139;
 Matches 705; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

```

Oy 1747 caaacgtagccttcttactaatagccagacatcagaactgcaggttagtataag 1806
Db 712 CAAAGTGAAGGCGCTTGTACTAATAGCCAGACATCAGACTTGCAGTAGATGTTAATG 653
Oy 1807 agatgaactattcttgcgagcctccggaaatcctaataatgttaagatggagacact 1866
Db 652 AGATACCTATTATTCGACGCGCTCGGAATCCTAATATTGTAATGAGGGACACACTT 593
Oy 1867 gaattcttgaccatctatattagagccc-tctctgttaatgcatatatacttgcct 1925
Db 532 GCATTTGTGACCATTTCTATTGAGGCCCTTCTCTGTTTATCATATTATCTTGTGCTT 533
Oy 1936 ttaactgttgaaatcctaattctaacctaaagtgctgccttagtactttcttgcgcct 1985
Db 533 TTAACCTGGAATCTATTCTTAACCTAAAGGTGCTGCCCTAGTACTTTCTTTGCGCCT 473
Oy 1986 ctgcgcctcttttcttcttccaaagcaactctgagccatgagcagcgaacaaactga 2045
Db 472 CTGCTGCTCTTTTCTCTTCCAAACAGCAACTCTGAGGCCATGAGCCAAACCTAGA 413
Oy 2046 ggtactgtccaccctgcgtctcatalaaggaaggaaggtccatcccttgcgtgcagag 2105
Db 412 GGTACTGCTCCACCTCCTCTCTCATTAAGGAAGGAGGAGGCTCATCCCTTGATTCGAGAG 353
Oy 2106 ggaaggaagatggtgtggaagcctcgcagagagagatagacatgacttgcacaacat 2165
Db 352 GGAGAGGAGAGATGTTGTGAGGCGCTCGAGAGACAGATAGACATGAGCTTTGACAAACAT 293
Oy 2166 ctgtgagctctctctgtctttagaataagcatgtaacattcttcatcccttattcc 2225
Db 232 CTGTAGGCTCTCTCTCTTGAATAGCATGATCATCTTTATTCATTCCTTATTTCC 233
Oy 2226 tacacaattgttttacttcttctgtgtgtgagactgtgagacacacacaatgtgt 2285
Db 232 TACATCATTTGTTTACTTCTTGTGCTGTGAGACTGAGTGAACACACAAAATGTGT 173
Oy 2286 tgacactgtatgctgcgagcagagacagactgacttgaacatggcgagagagcc 2345
Db 172 TGACACTGTGATGCGGAGGAGGAGGAGGAGCTACTGACTTTGAACTGGGAGAGAGGCC 113
Oy 2346 ctgagctcatcacaagcccaactcttcttccctccaagtacagtgacactctgtgctcat 2405
Db 112 CTGGATATCATCCACCCACCTCTTTTCCCTCCAGTACAGTAAACACTTCGTGCTCCCAT 53
Oy 2406 tggcagatgagcacttccctgcagccataactgactgtcttgcgaattcttc 2457
Db 52 TGGCAGATGGCGACTTCCTCGACCCATACATGATGCTTTGTGATTTGCCCC 1

RESULT 3
BE727698 632 bp mRNA EST 15-SEP-2000
LOCUS 601564388F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3834093 5',
DEFINITION mRNA sequence.
ACCESSION BE727698
VERSION BE727698.1 GI:10141690
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 632)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCPD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

```

```

FEATURES
    source
        1..632
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3834093"
            /clone_1id="NIH_MGC_20"
            /tissue_type="melanotic melanoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: skin; Vector: pORF7; Site: 1: XhoI; Site: 2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5'
            adaptor: GGCACGAG(G). Size-selected >500bp for average
            insert size 1.8kb. Library constructed by Ling Hong in
            the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 210 a 168 c 140 g 114 t
ORIGIN

Query Match 21.3%; Score 606; DB 135; Length 632;
Best Local Similarity 100.0%; Pred. No. 7.6e-121;
Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 475 aacttccttgatgaaggtttctgcagacagagactatagaagaacagagagaaga 534
Db 1 AACTCTCTTGATGAGGTTCTCTGCAGCAGGAACTAATAGAAACGACGAGAGAGAA 60
Oy 535 gaactgaagaactgaagaatcacagaataaccacgaagagtgtaattctcaagag 594
Db 61 GAACGGAAGAAGACTAAGGAATACGAATAAACCTCAAGAGGTTGGAATTTCTCAAG 120
Oy 595 aacaaagaagatgtaagaagaactgactgtgaagcctatagaacaacaaagaagtc 654
Db 121 AACCAAGAAGAGTGAAGAAAGAACTGACTGTGAAGCTTATAGAAACCAAGCAAGTTC 180
Oy 655 tcccaagcgaagctgtgtgcagagagctgtgaagcctagaagctagaagaggaagc 714
Db 181 TCCAGGCGAAGCTTTGTGGCAGAGCTGTGAAGCTAATAGAGCTAGAGAGTGGCAACGT 240
Oy 715 gtgaagaagactgaacccggaacccctgagcagaatgaacgaagaagcccatccctgc 774
Db 241 GTGAAGAAGACTGAACCGGAGCCCTGAGCCAGATGACAAAGAAATCAGACCCCTCATCTGC 300
Oy 775 aagttctcgaaacaacctccctgagtgagccctccatccactgcccctctgtgcagta 834
Db 301 AAGTCTCTCGAACAACCTCCCTGTGAGTGCGCCCTCATCCACTGCCCTCTGCTGCAGTA 360
Oy 835 tttatcggcatcctccagagcctgtgtgctactcttggagagcagactcgaagtcagc 894
Db 361 TGTATCGGATCTCTCCAGGCGCTGTGGGTCTACTCTTGGAGACGAGCATCCGATCCAGC 420
Oy 895 tcaagaacagaaagcaccatcaatgcacacggaagaatgtctctccatctccgaacc 954
Db 421 TCAGACAGGAAAGGACCATCAATGCACCGGAAGAAGATTGTCTCCATGCTTCGAAGC 480
Oy 955 aacacctccctcgaagcccccctagttctcgtgctccctacaaggaagctcctcccaag 1014
Db 481 AACACCTTCTCGAAGGCCCTCAGTTTCTCGTCCCTCAACAGGAGAGCTCCCTCCCAAG 540
Oy 1015 gtagtggagacgttcatgctgagcctatagagatattgcccctcaaaaaaaactcttt 1074
Db 541 GTAGTGGAGACCGTTCAATGCTGCTATAGGCAATTATGTCTCCCAAAAAAAACCTCTTG 600
Oy 1075 cctgca 1080
Db 601 CCTGCA 606

RESULT 4

```

found through the I.M.A.G.E. Consortium/LNL; at: image.lnl.gov
 Plate: L1CM513 row: 0 column: 22
 High quality sequence stop: 620.

AM821142
 LOCUS AM821142 614 bp mRNA EST 17-MAY-2000
 DEFINITION PM2-ST0303-030200-005-d12 ST0303 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AM821142
 VERSION AM821142.1 GI:7914136
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 614)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM2-ST0303-030200-005-d12&t=3-2000-02-03&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 582.
 Location/Qualifiers
 1. 614
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ST0303"
 /dev_stage="Adult"
 /note="Organ: stomach; Vector: puc18; Site_1: Sma1;
 Site_2: Sma1; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 166 a 185 c 135 g 128 t
 ORIGIN

Query Match 20.9%; Score 596.2; DB 95; Length 614;
 Best Local Similarity 99.5%; Pred. No. 1e-118;
 Matches 598; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 584 ttctcaagaagaagaagaagtgtgaagaagaactgtgtgaagcctatagaacca 643
 |||||||
 DB 14 ttctcaagaagaagaagaagtgtgaagaagaactgtgtgaagcctatagaacca 73
 |||||||
 QY 644 agaacaaagtctcccaaggagaagctgttgcaagagctgtgaagcctatagaacca 703
 |||||||
 DB 74 AGAACAAAGTCTCCCAAGGCAAGCTGTGGCAGAGAGCTGTGAAGCATTAAGAGCTCAGAGA 133
 |||||||
 QY 704 gtggcaacagtgtaagaagactgaacccgagcctgagccagatgaagaatcaagagc 753
 |||||||
 DB 134 GTGGCAACAGTGTGGAAGAAAGACTGAACCGGAGCCTGTGAGCAGATTAAGAGCTCAGAGA 193
 |||||||
 QY 764 cctcatctcgaagctctctggaaagacccctcctgagtggtgcccctcatcactgacct 823
 |||||||
 DB 194 CCTCATCTCTCAAGTCTCTGTGGAACACCTCCCTGTGAGTGAGCCCTCCATCCACTGCCCCCT 253
 |||||||
 QY 824 ctgctgagatgtatgcatactctcccaagcctgtgtgcttactctctggagagcagcact 883
 |||||||

|||||
 DB 254 CTGCTGAGATATGATGGCATCTCTCCAGAGCCGTGGCTACTGTGGAGAGGCACT 313
 |||||||
 QY 884 ccgagttccagctcagacagcagaagcaacatcaatgaccacggaaagaatgtctctcca 943
 |||||||
 DB 314 CCGAGTCCAGCTCAGACAGGAGGACCAATCAATGCCACCGGAAGATGTCTCTCA 373
 |||||||
 QY 944 ttctcgaacaacacctctctgagggcccccagtttccgtccctacacggagct 1003
 |||||||
 DB 374 TCTTCCAGCAACCAACCTCTCTGAGGCCCCCTAGTTTCCTCCCTACACAGGAGCT 433
 |||||||
 QY 1004 cctcccaaggtatagcagcagcgttcatgtcgtcctatagcattatgtctcccaaaaa 1063
 |||||||
 DB 434 CTTCCCAAGGATGATGAGACCGTTATGCTGCTTATGACATTAATGCTCCCAAAAA 493
 |||||||
 QY 1064 aaactcctgtcgtcatcctgtgtatacaatgacattttaaccaatccaatcaaaa 1123
 |||||||
 DB 494 AAATCCTTGTGCGTCGATCTGTGTACATGACATTTTAACCAATCAATCTAATAA 553
 |||||||
 QY 1124 tgtgcaagatccacctgtgtgcccgaatcgtgtgttctctcttactcaactgcag 1183
 |||||||
 DB 554 TGTGCCAGAAATCCACCTGTGAGCCGAACTGTTGTGCTCTTCTACTCCATGACAG 613
 |||||||
 QY 1184 a 1184
 |||||||
 DB 614 A 614
 |||||||

RESULT 5
 BE387718
 LOCUS BE387718 638 bp mRNA EST 21-JUL-2000
 DEFINITION 601287377F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3621618 5',
 mRNA sequence.
 ACCESSION BE387718
 BE387718.1 GI:9333083
 VERSION BE387718
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 638)
 NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 UNPUBLISHED (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM295 row: j column: 19
 High quality sequence stop: 634.
 Location/Qualifiers
 1. 638
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3621618"
 /clone_lib="NIH_MGC_44"
 /tissue_type="endometrium, adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the Laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 160 a 159 c 141 g 178 t
 ORIGIN

Query Match 20.9%; Score 596.2; DB 107; Length 638;
 Best Local Similarity 99.3%; Pred. No. 1e-118;
 Matches 609; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

OY 1013 gggtagatcgacgcttcacgtctgcctataggcatltagtccctcaaaaaaactcctt 1072
    |||||||
DB 1 GGGTAGATCGGACGCTTCATGCTGCTATAGGCAATATATGCTCCCAAAAAAATGCTCTT 60
OY 1073 tggcgcgtccgtgttacaacatgacattttaaccaatccaatcaaaatggtccaga 1132
    |||||||
DB 61 TGGCTGCATCCGTGTGTACACATGACATTTTAAACCAATCAATCTAAAAATGTGCCAGA 120
OY 1133 atccacgtgtgcccgaatcgtgtgtgtctctctcttctactccaactgcagatgaccaa 1192
    |||||||
DB 121 ATCCACGCTGTGGCCCGAATCGTGTGTGCTCTTCTTCTACTCCATGCAATGACAA 180
OY 1193 cctgtccgcgtgcacacttctcactgatatgtggagagaggaagcccaagccgaagt 1252
    |||||||
DB 181 CCGTCCGCGCTGCCACTTCTCCTACTGATATTGGAGAGGAGGCGCAAGCCGCAAGTT 240
OY 1253 caacataaaatgccccagagagatagcgacccgctgctgtccaaaggttggtttta 1312
    |||||||
DB 241 CCACATAAAATGCCAGAGAGATAGGCAACGGCTGGCTTGGCAAGGCTTTGGTTTAA 300
OY 1313 ttgcttctgttttcttcttccgcagacacaaagaagtaaggcagttatggacag 1372
    |||||||
DB 301 TTGCTTCTGTTTCTTTTCTTCCGACACACAAAGAACTAAGGCGAGTATTGGACAG 360
OY 1373 tgttatttaaacattctatgttaaatgaatgtgtgtgtgtgttctactgcatgtggagc 1432
    |||||||
DB 361 TGTATTTTAAACATCTATGTAATGAATGTGTGTTGTCTTACTGATGTCATTTGGAGAC 420
OY 1433 atgcggggggaagaagacccgacccaggttaagatggaagcccttccctggaaactaacg 1492
    |||||||
DB 421 ATGGGGGGAGAGAACTGACCCAGGTATGTAATGAGACCCCTTCCCTGGAACTAACCAG 480
OY 1493 tccctgattgtgtgtgactaaagtaaatgaataaaccacatcgtc-99999tgcactt 1551
    |||||||
DB 481 TCCCTGATGTGTGTGACAACTAAGATGAATGAATGCCATGTGCTGGGGGTGTGCACCT 540
OY 1552 cacactggcatgcatgtgtgaaagcttccatacccttggccatccctctctctctct 1611
    |||||||
DB 541 CACACTCGGCAATGCAATTTGAAAGCTTTCCATACCCCTTGGCATTCCCTCTCTCTCT 600
OY 1612 ctccaccccaatt 1624
    |||||||
DB 601 CTCACACCCATT 613
  
```

RESULT 6
 BE732904 917 bp mRNA EST 15-SEP-2000
 LOCUS 601567847F1 NIH.MGC_21 Homo sapiens cDNA clone IMAGE:384286 5',
 DEFINITION mRNA sequence.
 BE732904
 VERSION BE732904.1 GI:10146896
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 917)
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNL at: image.ilnl.gov
 Plate: L1CM536 row: m column: 11
 High quality sequence stop: 841.

FEATURES
 source Location/Qualifiers

```

1. 917
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:384286"
/clone_1db="NIH.MGC.21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: placenta; Vector: pOTB7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(C). Size selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the Laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 294 a 192 c 251 g 180 t
 ORIGIN

Query Match 20.8%; Score 594.2; DB 135; Length 917;
 Best Local Similarity 92.0%; Pred. No. 2.9e-118;
 Matches 717; Conservative 0; Mismatches 48; Indels 14; Gaps 8;

```

OY 201 tggtagttcattatgtagtgagggatgtagtgaacttattcaaaaagagttgt 260
    |||||||
DB 98 TGGTGTTCATATTGATGAGGAGGGGATGATGTAACCTATTATCAAAAAGAG-TTGT 156
OY 261 gctcagcagacactagatgtaacgcgcgaagagggcgcaagaagaatgaggaagttcg 320
    |||||||
DB 157 GTCTAGGCGAGACTAGATGAACGGCGCAAAAGAGGCAAGAAATGGAGAAAGTTGC 216
OY 321 aaacctgaagaatccagaagaatgctccagaagaggtttatgacccctcgtactataga 380
    |||||||
DB 217 AAACCTGAGATCCAGAAAGATGTCCAGAGAGGTTTATGACCCCTGATCTATATAGA 276
OY 381 aagcttacaggaacagaaggaagacagaggaagtaagtaagtaagtaagtaagtaag 440
    |||||||
DB 277 AAGGCTACAGGAACAGAGGACAGAGAGCAGAGATGACGAGGAACGTTCAAAATTCAA 336
OY 441 aaacatgtgaagagcttagatgaatgtagacacacttcccttgatgagtttcgcaga 500
    |||||||
DB 337 AATCATGATAGAGGCTTAGATGAATGATGACACAACTTCTTGATGAGGTTTCTGACA 396
OY 501 gcaagaactaataagaagaacgaagaagaagaagaactgaagaactgaagaactag 560
    |||||||
DB 397 GCAGGACTAATATAGAAAGCAAGCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 456
OY 561 aataacctcaagaaggttggaattctcaagaagaagaagaagaagtggaagaact 620
    |||||||
DB 457 AAATAAAGCTCAAGAGGTTGGATTTCTCAAGAGAAAGAGGAAGTGGAAAGAAACT 516
OY 621 gactgtgaagcctataagaagaacaaagaagttcccaaggaggaagcgtgtggcaggagc 680
    |||||||
DB 517 GACTGTGAAGCCTATAGAAACCAAGACAAAGTTCTCCAGGCGGAGAGCTGTTGGCAGAGC 576
OY 681 tctgaagataaagagctcaagagatgagcaacagtgtaaaagaactgaacaggacagcga 740
    |||||||
DB 577 TGTGAAGCATATAGAGCTCAGAGAGGCGCAAGCTGTGAAAAGACGTGAACCGCA-CTGA 635
OY 741 gccagatgaacaaagaaatcaagag-ccctcatcctgc-aagtcctcgtgaagaacactccctg 798
    |||||||
DB 636 GCCAGATGACAAAGATCAAGAGCCCTCATCTCTCAAGAGTCTCTGGAACACCTCCCTG 695
OY 799 agtggccctcatccactgcccctctgctgagatgatgatgatgatccctccaggctc 858
    |||||||
DB 696 AGTGG-CCCTCCATCCACTGCCCTCTCTGTCGATGATATATCG---ATCTCCAGAGCT 750
  
```


| FEATURES | source |
|--|---|
| BASE COUNT | 218 a 117 c 198 g 166 t |
| ORIGIN | 1..699 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3849620" /clone_lib="NIH.MGC.65" /tissue.type="adenocarcinoma" /lab_host="DH10B (phage-resistant)" /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies." |
| Query Match | 20.6%; Score 588.2; DB 136; Length 699; |
| Best Local Similarity | 95.4%; Pred. No. 5,5e-117; |
| Matches 669; Conservative 1; Mismatches 24; Indels 7; Gaps 6 | |
| RESULT 7 | |
| BE869679 | 699 bp mRNA EST 27-SEP-2000 |
| LOCUS | 601445783p1.NIH.MGC_65 Homo sapiens cDNA clone IMAGE:3849620 5', |
| DEFINITION | mRNA sequence. |
| ACCESSION | BE869679 |
| VERSION | BE869679.1 |
| KEYWORDS | EST. |
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 699) NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L10CM554 row: f column: 21 High quality sequence stop: 690. |
| COMMENT | Location/Qualifiers |
| JOURNAL | |
| TITLE | |
| AUTHORS | |
| DB | 919 gccaccggaagatgtctctccatcttcacgcgaaccaacacctctcgagcccta 977 11 111 111 111 111 111 111 111 111 111 111 111 111 111 111 806 gcaccggaagatgtctctccatcttcacgcgaaccaacacctctcgagcccta 864 |
| Db | 751 gggtgctaatcttggagca--ggatccgagctccagctgacagggag--ccatcaat 805 |
| Qy | 8 gaagggctcaaaaagaatacaagaagagcattcttttttttttttttttttttt 67 1 GAAGGGTCAAAAAGAAATACAAAAGCAAGGCTATTTCTTTTCTTTCTTTCTTCA 60 |
| Qy | 68 tcmctccttcctctgttcttcttcttcttcttcttcttcttcttcttcttcttct 127 11 127 |
| Db | 61 ttccttcccttcccttcttcttcttcttcttcttcttcttcttcttcttcttct 120 121 gcgcctctgcgcgtgcgcgttgaggcgccgcgcgcgcgcgcgcgcgcgcgcgc 180 |
| Qy | 128 gcgcctctgcgcgttgaggcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 187 121 gcgcctctgcgcgttgaggcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 180 |
| Qy | 188 cgcgcgcgtatagttgtgtttcattatagatgagggggagatgagttaaccttatca 247 181 cgcgcgcgtatagttgtgtttcattatagatgagggggagatgagttaaccttatca 240 |

| REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
|------------|---|---------------------------------|---------|---------|
| 248 | aaaagagttgtgtgtcgtgagcagacactgata | ggaagcgccaaaggaaggaaggaat | 307 | |
| 241 | aaaaagagttgtgtgtcgtgagcagacactgata | ggaagcgccaaaggaaggaaggaat | 300 | |
| 308 | ggagagaagttcgaagaaccttgaaagatccaa | gaagaatgctccagaagagtttatgaccctc | 367 | |
| 301 | gggagaaaatttcgaagaaccttgaaagatccaa | gaagaatgctccagaagagtttatgaccctc | 360 | |
| 368 | gattctctatatgaaggctctacagacagaaaga | agacagacagcagagagttacaggaac | 427 | |
| 361 | gattctctatatgaaggctctacagacagaaaga | agacagacagcagagagttacaggaac | 420 | |
| 428 | agttcaaatccaaaaacatggtgaagagccttga | tatgatagtaagatgagaccacttcctgatg | 487 | |
| 421 | agttcaaatccaaaaacatggtgaagagccttga | tatgatagtaagatgagaccacttcctgatg | 480 | |
| 488 | aggtctctcgaagaagaagacactataagaaga | caagaagaagagaagaagacttgaagaac | 547 | |
| 481 | aggtctctcgaagaagaagacactataagaaga | caagaagaagagaagaagacttgaagaac | 548 | |
| 539 | tgaaggaattacagaaataaactcctcaagaaga | aggttggaaatttcacagagaacaagaggag | 598 | |
| 548 | tgaaggaattacagaaataaactcctcaagaaga | aggttggaaatttcacagagaacaagaggag | 607 | |
| 608 | tggaaaagaagactgtgtaagagcctatagaaca | agaagaagaagttctccagaggagac | 667 | |
| 599 | gggaaaagaaa-tgactgtgaaaggcttatgaaac | -aaagaaaagttctccacagcgcaacg | 655 | |
| 668 | tgt-tggcagagagctgtgaag-cataaagagctc | agaagagtg | 706 | |
| 656 | tgtggggcagagagctgtgtgaagccatragagct | ctcagagagtg | 696 | |
| RESULT | 8 | | | |
| LOCUS | BE392702 | 664 bp | mRNA | EST |
| DEFINITION | 601307895f1 NIH_MGC_44 Homo sapiens | CDNA clone IMAGE:3626302 | 5' | |
| ACCESSION | BE392702 | mRNA sequence. | | |
| VERSION | BE392702.1 | GI:9338166 | | |
| KEYWORDS | EST. | | | |
| SOURCE | human. | | | |
| ORGANISM | Homo sapiens | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | |
| | 1 (bases 1 to 664) | | | |
| | NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ | | | |
| | National Institutes of Health, Mammalian Gene Collection (MGC) | | | |
| | Unpublished (1999) | | | |
| | Contact: Robert Strausberg, Ph.D. | | | |
| | Tel: (301) 496-1550 | | | |
| | Email: Robert.Strausberg@nih.gov | | | |
| | Tissue Procurement: ATCC | | | |
| | CDNA Library Preparation: Ling Hong/Rubin Laboratory | | | |
| | CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL) | | | |
| | DNA Sequencing by: Incyte Genomics, Inc. | | | |
| | Clone distribution: MGC clone distribution information can be | | | |
| | found through the I.M.A.G.E. Consortium/ULNL at: | | | |
| | http://image.llnl.gov | | | |
| | Plate: L1CM307 row: m column: 23 | | | |
| | High quality sequence stop: 631. | | | |
| FEATURES | | | | |
| SOURCE | | | | |
| | 1..664 | | | |
| | /organism="Homo sapiens" | | | |
| | /db_xref="taxon:9606" | | | |
| | /clone="IMAGE:3626302" | | | |
| | /clone_lib="NIH_MGC_44" | | | |
| | /tissue_type="endometrium, adenocarcinoma cell line" | | | |
| | /lab_host="DH10B (phage-resistant)" | | | |
| | /note="Organ: uterus; Vector: pOTB7; Site:1; XhoI: Site_2" | | | |
| | Ecot1: CDNA made by oligo-dt priming. Directionally | | | |
| | cloned into EcotRI/XhoI sites using the following | | | |
| | adaptor: GGCACGAG(G). Library constructed by Lin | | | |

in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 158 a 176 c 148 g 182 t

Query Match 20.5%; Score 583.4; DB 107; Length 664;
Best Local Similarity 98.4%; Pred. No. 5.9e-116;
Matches 631; Conservative 0; Mismatches 6; Indels 4; Gaps 4;

914 tcaatgccacccggaagaattgtctctccatcttcgaacacacctctctcgagccc 973
1 tcaatgccacccggaagaattgtctctccatcttcgaacacacctctctcgagccc 60
974 cctagttctccgtccctcacaagaggagctctcccaagggtatagtcgacgttcacg 1033
61 cctagttctccgtccctcacaagaggagctctcccaagggtatagtcgacgttcacg 120
1034 ctgcctataggcatatgtccctcaaaaaaactccttgcctgcctgcctgcctgcctgc 1093
121 ctgcctataggcatatgtccctcaaaaaaactccttgcctgcctgcctgcctgcctgc 180
1094 atgacatttttaacccaatccaatcaaaatgtccgaagatccacgtgtgcccgaatcg 1153
181 atgacatttttaacccaatccaatcaaaatgtccgaagatccacgtgtgcccgaatcg 240
1154 tctttgtct 1213
241 tctttgtct 300
1214 tcaactgatattggaag 1273
301 tcaactgatattggaag 360
1274 aataggaacccgagcgtgcttgcacaaagggttgggtttatgtcttcttcttcttct 1333
361 aataggaacccgagcgtgcttgcacaaagggttgggtttatgtcttcttcttcttct 420
1334 cccgac 1393
421 cccgac 480
1394 taa-atgaatgtgttctgttctgttctgttctgttctgttctgttctgttctgttctgt 1452
481 taaatgaatgtgttctgttctgttctgttctgttctgttctgttctgttctgttctgt 540
1453 cccgaggtatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1511
541 cccgaggtatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 600
1512 aagtaagaatataaaccatctgtgaggtgtgtcattc 1552
601 TAGT-AAGATGATATAA-CCCATCTGCTGGGGGTGCTCTTTC 639

RESULT 9
BE180639/c
LOCUS BE180639 631 bp mRNA EST 22-JUN-2000
DEFINITION RC3-HR0625-040500-022-d11 HT0625 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE180639
VERSION BE180639.1 GI:8659815
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordino, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

TITLE
JOURNAL
MEDLINE
COMMENT
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=612-RC3-HR0625-040
500-022-d11&f3=2000-05-04&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 5
High quality sequence stop: 630.
Location/Qualifiers
1. 631

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="HT0625"
/dev_stage="Adult"
/note="Organ: head,neck; Vector: puc18; Site:1: Sma1;
Site:2: Sma1: A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 127 a 145 c 182 g 177 t

Query Match 20.4%; Score 581.8; DB 105; Length 631;
Best Local Similarity 98.7%; Pred. No. 1.3e-115;
Matches 597; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

542 aagaactgaagaatcacagaataaccccaagaaggttgaattctcagaagaacaaga 601
631 AAGAACTGAGGAGATGAGAAATTAACCTCAAGAGGTTGGAATTTCTCAAGAGGACAGA 572
602 aggaagtgaagaagaactgaactgtgaagcctatagaacccaagaagaacttccagag 661
571 AGGAAGTGGAAAGAAATGACTGTGAAGCCTATGAAACCAAGAAAGTTCTCCACAG 512
662 cgaagctgttgcagagagctgtgaagcataagagctcagagagtgcacagtgtaaaa 721
511 CGAAGCTGTGGCAGAGACTGTGAAGCATTAAGAGCTCAGAGAGTGCAACAGTGAAGA 452
722 gactgaacccggaacccctgaagcagatgacagaatcaagaagacctcctcgaagcttc 781
451 GACTGAACCCGAGACCCCTGAGCAGATGAACCAAGAAAGAGGCCCTCATCTGCAAGTCTC 392
782 tcggaacacacctccctgtgagtgagccctcatcacatcagtcgcccctgtcagtatgacg 841
391 TCGGAACACCTCCTCTGAGTGGCCCCCTCATCTCAGCTGCCCCCTCTGCTGAGTATATCG 332
842 gcatctcccaagcctgtggtgctactcttgcagagcagcagactcgaagcagcagcagc 901
331 GCATCTCCCAAGCCTGTGGTGTGCTCTGTGGAGAGAGGAGCTCCGAGTCCAGACTCAGACA 272
902 gcgaagcagcatcaatgacacccggaagaatgtctctccatctccgaacacacacct 961
271 GCGAAGCAGCATCAATGCCACCGGAAGATGTCTCTCCATCTTCCGAACCAACACACT 212
962 tctctgagagccctcagttctcgtccctcagaagagagctcctcccaagggtgagctc 1021
211 TCTCTGAGAGCCCTCAGTTCTCTGCTCTCTACACAGGAGGACTCTCTCCCAAGGATGATC 152

QY 1022 ggaagcttcatgctcgtatagcattatgcccccaaaaaaacctcttgcctgcat 1081
|||||
Db 151 GGAGCGTTGATGCTGCGCTGATGCGATATATGCTTCCATAAAAAAAGCTTGTGCTGAT 92
QY 1082 ccgtgtgacaacatgacatttttaacccaatccaatcaaaatgctcagaatccactg 1141
|||||
Db 91 CCGTGTGACAAACATGATGATTTTAAACATGCAATCTAAATATGTCAG-ATCCACCTG 33
QY 1142 tggcc 1146
|||
Db 32 TGCCC 28

RESULT 10
BE736929 895 bp mRNA EST 15-SEP-2000
LOCUS BE736929
DEFINITION 601306864F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3641343 5',
mRNA sequence.
ACCESSION BE736929
VERSION BE736929.1 GI:10150832
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 895)
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLCM346 row: p column: 16
High quality sequence stop: 726.
Location/Qualifiers
1. 895
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3641343"
/clone_lib="NIH_MGC_39"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(d). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-CDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."

BASE COUNT 232 a 183 c 218 g 262 t
ORIGIN

Query Match 20.2%; Score 575.2; DB 135; Length 895;
Best Local Similarity 91.9%; Pred. No. 3.7e-114;
Matches 741; Conservative 0; Mismatches 48; Indels 17; Gaps 12;

QY 1170 ctactccatgcatgacacaaactgtccgctgcacttctactgataatggag 1229
|||||
Db 1 CTACTCCATGACATGACCAACCTGCTCCGCTG-CACCTTCCTCAGTATTA-TGGAG 58
QY 1230 gagggaagggccagcgaagtccactaaatgccccagaggaataggcagcgctg 1289
|||||
Db 59 GAGGGAGAGGCCAGCCGGAAGTTCCATAAATATGCCAGGAGATAGGCACCGGCTG 118

QY 1290 ctggcaagaggttgggttatttcttcttcttcttccagacgaaga 1349
|||
Db 119 CTGCAAGAGG---TTGGGTTTATGCTTCTGTTTCTTCTTCCGACAGCAAGA 175
QY 1350 agtaaggcagattatggaaggtgtatttaaacattcatgttaaatgltgt 1409
|||||
Db 176 AGTAAGGAGGAGTTA-TGACAGAGGTTATTAACATTCTA-TGTAAATGAATGTG -G 231
QY 1410 ttgttctactgatttggagcatggyggggaaggaactgcccagtaagaaatg 1469
|||||
Db 232 TTGGTTCTACTGATGTTGTGAGCATGCGGGGAGAGAACTGACCAGGTAATGAATG 291
QY 1470 agccctccctggaactcaacagtccttgaatgtgtgtgactaagtaagataaac 1529
|||||
Db 292 AGCCCTTCCCTGAGACATCAACAGTCTTGTATGTTGTGCTAAGTAAGATATTAAC 351
QY 1530 ccactgtcgggggtgtcacttaacactggcatgtcatttgaagcttccataccct 1589
Db 352 CCATCTGCTGGGGGTGTCACCTTCACACTCGCATGTCATGTGAAGCTTCCATACCTT 411
QY 1590 ggcattccctctctctctctctcccaaccattatgcaggaaggaactgtaacaag 1649
|||||
Db 412 GGCCATTCCCTCTCTCTCTCTCCCAACCCATTATGCGAGAGGAGCTG-TAACAG 470
QY 1650 aacgcttccatccaaaccttctctctgctgggaattatattgttlttgaa 1709
|||||
Db 471 AACGCTTCCATCTCAACCTTCTCTGCTGGGAATTTATATGTT-GTTTGAAG 538
QY 1710 taaaggaattagtttaagattcctaaatttaagaaacaaagtagcctgttactaa 1769
Db 529 TAAAGGATTTAGTTAAGATTCAATTTAGAGAAACAACGTAGGCTTGTACTTA 588
QY 1770 tagcagaacatcagaactcaggttaggtatgttaatgagatgactaatttcgagctc 1829
Db 589 TACCCAGACATCAAGACTG-AGGTAGGTATGTAATGAGATGACTTATTTCTGCGAGCTC 647
QY 1830 ctggaatcccaatattgtaataga--gtggagacacctgcataatgtgacattatc 1887
Db 648 CTGGAATCTTAATATTGTAATATACGTGGCAACAACCTTGCTATGTTGACATTTCTATT 707
QY 1888 gaagccctctctgtttaaagcatatatacttgcatttaactgtgaactatctta 1947
|||||
Db 708 GAGGCGCTTCTGTTTATGCT-TATATACTTGTGCTTT-ACGGGGGAGTATTTCTTA 765
QY 1948 acctaaaggtgcgcctagtaactt 1973
Db 766 CTTAAGTGTGGCTCTAGTAATTTTTT 791

RESULT 11
AM327476 568 bp mRNA EST 28-JAN-2000
LOCUS AM327476
DEFINITION dq03909.x1 NIH_MGC_2 Homo sapiens cDNA clone IMAGE:284800 5', mRNA
sequence.
ACCESSION AM327476
VERSION AM327476.1 GI:6797971
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 568)
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Edge Biosystems
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center (NISC)
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bbrp/image/image.html

Plate: L1C00028 row: N column: 17
Seq primer: -21M13 forward primer (ABI).

FEATURES

Source

Location/Qualifiers
1..568

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2846800"
/clone_lib="NIH MGC 2"
/tissue_type="T cell leukemia"
/cell_line="MGC2"
/note="Organ: Blood; Vector: pOTB7a; Library prepared by Edge Biosystems."

BASE COUNT 164 a 165 c 123 g 116 t
ORIGIN

Query Match 19.8%; Score 564.8; DB 89; Length 568;
Best Local Similarity 99.6%; Pred. No. 6.1e-112;
Matches 566; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 558 cagaataactcaagaaggttggaatttctcaagaagaagaagaagtgaaagaa 617
DB 1 CAGAAATACCTCAAGAAAGTTGGAATTTCTCAAGAAACAAAGAAAGTGAAGAA 60
OY 618 actgacttgaaagcctatagaaccaagaagaagctcccgagcgaagctgtgcaag 677
DB 61 ACTGACTGTGATCTCTATAGAAACCAAGACAGATTTCTCCGCGGAAGCTTTGGCAG 120
OY 678 agctgtgaagcaataagaagctcagaagatgtgcaacagtttgaagaactgaacccgagacc 737
DB 121 AGCTGTGAAGATTAAGACTCTGAGAGAGTGGCAACAGTGTGAAGAACTGAAACCGGACCC 180
OY 738 tgaagcagaatgaagaagaatcaagaagcctcaatcctcgaagctcgcgaagaacatcctcct 797
DB 181 TGAGCCAGATGACAAAGAAATCAAGAGCCCTCATCTGCAAGTCTCTGGAAACACTCTCCT 240
OY 798 gagtggccctccatccactgcccctctgtcgaatgtatcgtacatccctccagagct 857
DB 241 GAGTGGCCCTCCATCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
OY 858 ggggtgctactcttgagagcagcagctccgagctcagctcagaacagcgaagcacaatcaa 917
DB 301 GGGTGCCTACTCTGGAGCAGGAGCTCCGAGTCCAGCTCAGACAGCAAGCAGCATCA 360
OY 918 tggcaccggaagaagatgtgtctcctcctcatcttcgaaacaaacccctcccgagccctccta 977
DB 361 TGCCACCGGAAAGATTGTCTCTCCATCTTCGGAACCAACACTTCTCGAGGCCCTCTA 420
OY 978 gttctccgtccctacacagagagctcctcccaaggttagatcgagccgttcatgtctgc 1037
DB 421 GTTCTCGCTCCCTACACAGGAGCTCTCCCAAGGGTAAATCGGACCTTCAATGCTGC 480
OY 1038 ctatagcattatgtccctcaaaaaaaactccttgccttgatccctgtgtgtaacaatga 1097
DB 481 CTATAGGCAATTATGTCTCCCAAAAAAACTCTTGGCTGATCTGTGTACACATGA 540
OY 1098 cattttaaccaatcattcaaaaatg 1125
DB 541 CATTTTAAACCAATCAATCTAAATATG 568

RESULT 12
AM592821/c 554 bp mRNA EST 22-MAR-2000

LOCUS AM592821 554 bp mRNA EST 22-MAR-2000

DEFINITION hg6601.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone

IMAGE:2944800 3', mRNA sequence.

ACCESSION AM592821

VERSION AM592821.1 GI:7280012

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 554)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.lnl.gov) for further information.
Seq primer: -40UP from Glibco
High quality sequence stop: 438.

FEATURES

Source

Location/Qualifiers
1..554

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2944800"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site: 1: Not I; Site: 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NDHL19W, testis NHT, and B-cell NCI-CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonalao."

BASE COUNT 151 a 122 c 135 g 146 t
ORIGIN

Query Match 19.4%; Score 552.4; DB 92; Length 554;
Best Local Similarity 99.8%; Pred. No. 2.9e-109;
Matches 553; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2249 tgggtgtgagactgagttagacacacacaaatgtgttgaacactgtgatccgcgagca 2308
DB 554 TGGGTGTGAGACTGAGTGAACACACACAAATGTTGTGACACTGTGATGCCGCGAGCA 495
OY 2309 gaggcagctactgacttgaacatggcagagagcccttgatcctacacagccaccc 2368
DB 494 GAGCAGCTACTGACTTTGAACATGGGCAAGAGGCCCTGGATCTCATCAAGCCACTCC 435
OY 2369 ttltccctccagtagacagctcgtgtgcccattgagcagatgagcacttccctga 2428
DB 434 TTTTCCCTTCACAGTAGAGTAGACACTGTGTCGCTGATGGCAGATGGGACTTCCTCA 375
OY 2429 ccataactgatgtcttgaattctccctccttccgaactactcgtgtcattatgtc 2488
DB 374 CCCATAGCTAGTGTGTAATTTCTCTCTCTTTCGAACTACTCTGTCTAATTTGT 315
OY 2489 ctgcaagatagggcagcagcactcactcactgaacaaagaacattagtaaaacttgc 2548
DB 314 CTGCCAGTAGGGCCATCAGCTCCATCTCTGACAAACAGACATTTAGGTAAACTTTGT 255
OY 2549 aggcactctgtctctgtcttcatgttctcgtgtatagctgtgtttaaagcat 2608
DB 254 AGGCACCTTCTGCTCTCTGCTTCAATTTCTCTCTGATGATGCTGTGTTATTACAGCAT 195
OY 2609 gtacccaaagcctcactatgttgaagaggcagcagcaggaacataagtcacatc 2668
DB 194 GTACCCAAAGACGCTCAATTTGTAAGAGGCGGCGAGACATCAAGCAACATCT 135
OY 2669 ttatgtgcatgactctaaagagcattactgtatcctcagagcctctgtgtgaaag 2728
DB 134 TTATGTGCATGACTTTAAGAGGCCATTAAGTATCTCATGAGCCCTTGTGATGGAAG 75

Thu Mar 1 08:56:57 2001

us-09-602-597-1.rst

Page 13

Qy 1627 tgcaggaaggactgtctaacaagaacgttc 1657
||||| | ||||||| ||||||| |||||||
Db 540 TGCACGGAAGACTGCTAACACAAGAAGCGTTC 570

```
Search completed: February 28, 2001, 23:43:20
Job time: 19386 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2001, 22:33:49 : Search time 115.04 Seconds
(Without alignments)
3992.584 Million cell updates/sec

Title: US-09-602-597-1

Perfect score: 2850
Sequence: 1 aagccctgaaggggtcaaaag.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NM:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/PCtUS.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|------------------|
| C 1 | 92 | 3.2 | 7218 | 1 | US-08-232-463-14 |
| 2 | 67.6 | 2.4 | 1307 | 2 | US-08-960-022-17 |
| 3 | 67.2 | 2.4 | 1641 | 1 | US-08-300-903A-8 |
| 4 | 67 | 2.4 | 2082 | 2 | US-08-785-310A-2 |
| 5 | 65.4 | 2.3 | 6671 | 1 | US-08-280-443-1 |
| 6 | 65.4 | 2.3 | 6671 | 1 | US-08-457-459-1 |
| 7 | 65.4 | 2.3 | 6671 | 1 | US-08-535-678-1 |
| 8 | 65.4 | 2.3 | 6671 | 4 | PCT-US95-02275-1 |
| 9 | 64.8 | 2.3 | 144 | 1 | US-08-702-344-26 |
| 10 | 64.8 | 2.3 | 1147 | 1 | US-08-665-716-1 |
| 11 | 64.8 | 2.3 | 1172 | 1 | US-07-945-288-9 |
| 12 | 64.8 | 2.3 | 1172 | 1 | US-08-462-831-9 |
| 13 | 64.8 | 2.3 | 1172 | 1 | US-08-461-809-9 |
| 14 | 64.8 | 2.3 | 1172 | 1 | US-08-461-441-9 |
| 15 | 64.8 | 2.3 | 1172 | 4 | PCT-US93-08518-9 |
| 16 | 64.8 | 2.2 | 1817 | 4 | US-08-473-981A-5 |
| 17 | 64 | 2.2 | 1817 | 2 | US-08-474-087-5 |
| 18 | 64 | 2.2 | 2381 | 1 | US-08-021-608D-9 |
| 19 | 64 | 2.2 | 2381 | 1 | US-08-726-160-9 |
| 20 | 64 | 2.2 | 2381 | 4 | PCT-US94-01782-9 |
| 21 | 64 | 2.2 | 2384 | 1 | US-08-021-608D-1 |
| 22 | 64 | 2.2 | 2384 | 1 | US-08-726-160-1 |
| 23 | 64 | 2.2 | 2384 | 4 | PCT-US94-01782-1 |
| 24 | 63.8 | 2.2 | 240 | 1 | US-08-628-417-6 |
| 25 | 63.8 | 2.2 | 5173 | 1 | US-08-242-677-1 |
| 26 | 63.6 | 2.2 | 140 | 1 | US-08-628-417-5 |
| 27 | 63.6 | 2.2 | 630 | 1 | US-08-185-414E-1 |
| 28 | 63.6 | 2.2 | 2233 | 1 | US-08-496-631-1 |

| | | | | | | |
|----|------|-----|------|---|-------------------|--------------------|
| 29 | 63.4 | 2.2 | 2447 | 2 | US-09-014-969-14 | Sequence 14, Appl |
| 30 | 63.4 | 2.2 | 2744 | 3 | US-09-071-101-1 | Sequence 1, Appl |
| 31 | 63.4 | 2.2 | 2744 | 3 | US-09-369-618-1 | Sequence 1, Appl |
| 32 | 63.4 | 2.2 | 2744 | 3 | US-09-369-617-1 | Sequence 1, Appl |
| 33 | 63.2 | 2.2 | 1882 | 2 | US-09-370-253-1 | Sequence 1, Appl |
| 34 | 63.2 | 2.2 | 2550 | 5 | 5258287-23 | Patent No. 5258287 |
| 35 | 63 | 2.2 | 253 | 2 | US-08-520-678A-25 | Sequence 25, Appl |
| 36 | 63 | 2.2 | 991 | 3 | US-08-924-747-25 | Sequence 25, Appl |
| 37 | 63 | 2.2 | 3527 | 2 | US-08-909-965C-7 | Sequence 7, Appl |
| 38 | 62.8 | 2.2 | 1066 | 1 | US-08-157-101A-4 | Sequence 4, Appl |
| 39 | 62.8 | 2.2 | 1098 | 3 | US-09-248-335-35 | Sequence 35, Appl |
| 40 | 62.8 | 2.2 | 1134 | 3 | US-09-248-335-29 | Sequence 29, Appl |
| 41 | 62.6 | 2.2 | 2852 | 3 | US-09-027-137-2 | Sequence 2, Appl |
| 42 | 62.4 | 2.2 | 1461 | 3 | US-08-722-126A-4 | Sequence 4, Appl |
| 43 | 62.4 | 2.2 | 1461 | 4 | PCT-US95-04258-4 | Sequence 4, Appl |
| 44 | 62.4 | 2.2 | 1738 | 2 | US-08-379-482A-2 | Sequence 2, Appl |
| 45 | 62.2 | 2.2 | 688 | 5 | 5498694-3 | Patent No. 5498694 |

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232.463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935.313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZgpc-F15
: US-08-232-463-14
: Query Match 3.2% Score 92; DB 1; Length 7218;

LOCATION: 3.839
US-08-300-903A-8

STREET: Spring House Corporate Cntr, P.O. Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA

Query Match 2.3%; Score 64.8; DB 1; Length 1147;
Best Local Similarity 68.2%; Pred. No. 5.8e-06;
Matches 90; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

[illegible]

```
Db      1125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1172  
|||||  
  
RESULT 12  
US-08-462-831-9  
Sequence 9, Application US/08462831  
Patent No. 5552142  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM  
DERMATOPHAGOIDES  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 STATE STREET, SUITE 510  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462.831  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/945,288  
FILING DATE: 10 SEPTEMBER 1992  
APPLICATION NUMBER: US 580,653  
FILING DATE: 11 SEPTEMBER 1990  
APPLICATION NUMBER: US 458,642  
FILING DATE: 13 FEBRUARY 1990  
ATTORNEY/AGENT INFORMATION:  
NAME: MANDRACOURAS, AMY E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1172 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..738  
US-08-462-831-9
```

```
Oy      2803 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2850  
|||  
Db      1125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1172  
Query Match          2.3%; Score 64.8; DB 1; Length 1172;  
Best Local Similarity 75.0%; Pred. No. 5.8e-06;  
Matches 81; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
```

```
RESULT 13  
US-08-461-809-9  
Sequence 9, Application US/08461809  
Patent No. 5770202
```

```

CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461.441
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/945.288
FILING DATE: 10 SEPTEMBER 1992
APPLICATION NUMBER: US 580.655
FILING DATE: 11 SEPTEMBER 1990
APPLICATION NUMBER: US 458.642
FILING DATE: 13 FEBRUARY 1990
ATTORNEY/AGENT INFORMATION:
NAME: MANDRACOURAS, AMY E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1172 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..738
US-08-461-441-9

Query Match      2.3% Score 64.8; DB 1; Length 1172;
Best Local Similarity 75.0%; Pred. No. 5.8e-06;
Matches 81; Conservative 0; Mismatches 27; Indels 0; Gaps 0.

QY 2743 ttgcagggtttaaaacatccattacaatgaagaactaatataaccggtcagagacaacaaa 2802
|| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1065 TTTTAGAAGAAATGAATTAATTCATTCTACCAAAATTTAAAAAAAATTTTTAAAAAAA 1124

QY 2803 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2850
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1172

RESULT 15
PCT-US93-08518-9
Sequence 9, Application PC/TUS9308518
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
NUMBER OF INVENTION: DERMATOPHAGOIDES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 17:05:06 ; Search time 16.5 seconds

(without alignments)
526.377 Million cell updates/sec

Title: US-09-602-597-2

Perfect score: 1306
Sequence: 1 MDGGDGNLIIKKRFVSEAE.....NATGKIYSSIFRTFTLEAP 254

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_36:.*
1: /SIDSL/gcgdata/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 779 | 59.6 | 168 | 21 | Human breast tumor |
| 2 | 143 | 10.9 | 1135 | 20 | Human acid sequence |
| 3 | 142.5 | 10.9 | 1178 | 18 | Mannose-1-phosphat |
| 4 | 141.5 | 10.8 | 562 | 16 | Leucocytotoxin prot. |
| 5 | 141 | 10.8 | 1239 | 20 | Human ZC1 protein. |
| 6 | 139 | 10.6 | 1233 | 20 | Mouse STRE20-relate |
| 7 | 133 | 10.2 | 1162 | 21 | HHV8 ORF 73 protei |
| 8 | 132.5 | 10.1 | 517 | 13 | 1-Caldesmon. Gall |
| 9 | 131.5 | 10.1 | 1898 | 20 | A human trichohyal |
| 10 | 129.5 | 9.9 | 1132 | 17 | Chicken leucocytos |
| 11 | 126 | 9.6 | 1360 | 21 | Human protein kina |
| 12 | 125 | 9.6 | 897 | 17 | Murine EGF recepto |

| | | | | | | |
|----|-------|-----|------|----|--------|---------------------|
| 13 | 125 | 9.6 | 897 | 19 | W47118 | Murine eps15 prote |
| 14 | 125 | 9.6 | 1297 | 20 | W94406 | Murine eps15 prote |
| 15 | 124.5 | 9.5 | 1297 | 20 | Y55932 | Human ZC2 protein. |
| 16 | 123 | 9.4 | 611 | 20 | Y29039 | T. gondii immunoge |
| 17 | 122 | 9.3 | 1227 | 20 | Y55965 | Full length human |
| 18 | 121.5 | 9.3 | 585 | 20 | W97757 | S-region transfer |
| 19 | 121.5 | 9.3 | 1118 | 19 | W82395 | Human UBP protein |
| 20 | 121 | 9.3 | 297 | 18 | W20698 | H. pylori cytoplas |
| 21 | 121 | 9.3 | 412 | 17 | W03626 | Human thymotrophi |
| 22 | 120.5 | 9.2 | 542 | 18 | W36450 | Human TULP1 protei |
| 23 | 120.5 | 9.2 | 1299 | 21 | Y58633 | Protein regulating |
| 24 | 120 | 9.2 | 360 | 17 | W03627 | Human follicle sti |
| 25 | 120 | 9.2 | 432 | 20 | W93954 | Human regulatory m |
| 26 | 119 | 9.1 | 185 | 20 | Y60129 | Human endometrium |
| 27 | 119 | 9.1 | 285 | 16 | R82689 | Shrimp tropomyosin |
| 28 | 118 | 9.0 | 480 | 20 | Y34653 | Amino acid sequenc |
| 29 | 118 | 9.0 | 1588 | 15 | R46605 | Malariol p16MP3 ep |
| 30 | 118 | 9.0 | 1663 | 15 | R46608 | Plasmodium falciipa |
| 31 | 116 | 8.9 | 411 | 20 | W68011 | Yeast immunophilin |
| 32 | 116 | 8.9 | 424 | 20 | Y49148 | Amino acid sequenc |
| 33 | 116 | 8.9 | 425 | 20 | Y49147 | Amino acid sequenc |
| 34 | 116 | 8.9 | 426 | 20 | Y49146 | Amino acid sequenc |
| 35 | 116 | 8.9 | 437 | 20 | Y49241 | N-terminal region |
| 36 | 116 | 8.9 | 437 | 20 | Y49244 | N-terminal region |
| 37 | 116 | 8.9 | 437 | 20 | Y32180 | N-terminal region |
| 38 | 116 | 8.9 | 437 | 20 | Y32183 | N-terminal choline |
| 39 | 116 | 8.9 | 439 | 20 | Y49240 | N-terminal region |
| 40 | 116 | 8.9 | 439 | 20 | Y32179 | N-terminal choline |
| 41 | 116 | 8.9 | 530 | 19 | W49042 | Human low density |
| 42 | 115.5 | 8.8 | 589 | 20 | Y43392 | PSPC alpha-helix c |
| 43 | 115.5 | 8.8 | 929 | 18 | W45593 | S. pneumoniae pneu |
| 44 | 115.5 | 8.8 | 929 | 18 | Y43384 | S. pneumoniae pspc |
| 45 | 115 | 8.8 | 223 | 20 | W96760 | A Helicobacter pyl |

ALIGNMENTS

| | |
|----------|---|
| RESULT 1 | |
| ID | Y48610 standard; Protein: 168 AA. |
| XX | |
| AC | Y48610; |
| XX | |
| DT | 08-DEC-1999 (first entry) |
| XX | |
| DE | Human breast tumour-associated protein 71. |
| XX | |
| KW | Expressed sequence tag; EST: human: breast: cancer: gene therapy; |
| KW | treatment; tumour; cytostatic; medicament. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | DE19813839-A1. |
| XX | |
| PD | 23-SEP-1999. |
| XX | |
| PF | 20-MAR-1998; 98DE-1013839. |
| XX | |
| PR | 20-MAR-1998; 98DE-1013839. |
| XX | |
| PA | (META-) METAGEN GES GENOMFORSCHUNG MBH. |
| XX | |
| PI | Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A; |
| XX | WPI: 1999-528981/45. |
| DR | N-PSDB; 233669. |
| XX | |
| PT | Human nucleic acid sequences and protein products from tumor breast |
| PT | tissue, useful for breast cancer therapy - |
| XX | |
| PS | Claim 22; 174; 188p; German. |
| XX | |

CC This invention describes novel human nucleic acid sequences from tumor
CC breast tissue which have cytosolic activity. The nucleic acid sequences
CC can be used to produce and isolate full-length gene sequences. They can
CC be used to express proteins, which can be used as tools to find an
CC activity against breast cancer. The sequences can be used in sense or
CC antisense form. They are especially useful for medicaments for gene
CC therapy to treat breast cancer. Y48540-Y48617 represent protein fragments
CC encoded by the expressed sequence tags described in the method of the
CC invention.

XX
SQ Sequence 168 AA;

Query Match 59.6%; Score 779; DB 20; Length 168;
Best Local Similarity 100.0%; Pred. No. 1,1e-59;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 MDGDDGNLIIRKRVSEAEIDERRRRROEEMKVRKPEDECEPEVYDPRSLYERLQ 60
13 mdggddgnlllkkrfvseaelderkrkrgeewekvrkpedeevydprslyerlq 72

DB 61 QKDRKQOEYEHOFKKNVRLDEETNPLDVGROQLIKORREELKELKEVRNLK 120
73 qkdrkqgeyeqikrkmvrgldetnpldevsrqgeljckgrreeelkeikyrnlk 132

DB 121 KVGISOENKKEVERKLTVPKPIETKNKFSQAK 151
133 kvgisgenkkekklvtpkpietknkfsqak 163

RESULT 2

ID Y68784
XX Y68784; standard; Protein; 1135 AA.

XX 16-MAY-2000 (first entry)

XX Amino acid sequence of a human phosphorylation effector PHSP-16.

KW Human: phosphorylation effector; PHSP; proliferative disorder;
KW Immune disorder; neuronal disorder.

OS Homo sapiens.

XX
FH Key Location/Qualifiers

FT Modified-site 9 /note- "potential phosphorylation site"

FT Modified-site 17 /note- "potential phosphorylation site"

FT Region 31..54 /note- "protein kinase signature sequence"

FT Modified-site 33 /note- "potential glycosylation site"

FT Modified-site 59 /note- "potential phosphorylation site"

FT Modified-site 59 /note- "potential phosphorylation site"

FT Modified-site 77 /note- "potential phosphorylation site"

FT Modified-site 112 /note- "potential phosphorylation site"

FT Modified-site 124 /note- "potential phosphorylation site"

FT Region 129..182 /note- "potential kinase signature sequence"

FT Region 149..161 /note- "protein kinase signature sequence"

FT Modified-site 187 /note- "potential phosphorylation site"

FT Active-site 190..200 /note- "potential phosphorylation site"

FT Active-site 214..236 /note- "tyrosine kinase catalytic site"

FT Modified-site 222 /note- "tyrosine kinase catalytic site"

FT Modified-site 255 /note- "potential phosphorylation site"

FT Modified-site 259 /note- "potential phosphorylation site"

FT Modified-site 264 /note- "potential phosphorylation site"

FT Modified-site 309 /note- "potential phosphorylation site"

FT Modified-site 319 /note- "potential phosphorylation site"

FT Modified-site 321 /note- "potential phosphorylation site"

FT Modified-site 323 /note- "potential phosphorylation site"

FT Modified-site 324 /note- "potential phosphorylation site"

FT Modified-site 326 /note- "potential phosphorylation site"

FT Modified-site 351 /note- "potential phosphorylation site"

FT Modified-site 467 /note- "potential phosphorylation site"

FT Modified-site 543 /note- "potential phosphorylation site"

FT Modified-site 550 /note- "potential phosphorylation site"

FT Modified-site 554 /note- "potential phosphorylation site"

FT Modified-site 570 /note- "potential glycosylation site"

FT Modified-site 572 /note- "potential phosphorylation site"

FT Modified-site 624 /note- "potential phosphorylation site"

FT Modified-site 625 /note- "potential phosphorylation site"

FT Modified-site 632 /note- "potential phosphorylation site"

FT Modified-site 681 /note- "potential phosphorylation site"

FT Modified-site 682 /note- "potential phosphorylation site"

FT Modified-site 688 /note- "potential phosphorylation site"

FT Modified-site 689 /note- "potential phosphorylation site"

FT Modified-site 706 /note- "potential phosphorylation site"

FT Modified-site 718 /note- "potential glycosylation site"

FT Modified-site 720 /note- "potential phosphorylation site"

FT Modified-site 726 /note- "potential phosphorylation site"

FT Modified-site 811 /note- "potential phosphorylation site"

FT Modified-site 815 /note- "potential phosphorylation site"

FT Domain 836..1115 /note- "potential phosphorylation site"

FT Modified-site 898 /note- "potential phosphorylation site"

FT Modified-site 931 /note- "potential phosphorylation site"

FT Modified-site 958 /note- "potential phosphorylation site"

FT Modified-site 978 /note- "potential phosphorylation site"

FT Modified-site 999 /note- "potential phosphorylation site"

| | | | |
|----|---|---------------------------------------|---|
| FT | Modified-site | 1012 | /note= "potential phosphorylation site" |
| PT | Modified-site | 1067 | /note= "potential glycosylation site" |
| FT | Modified-site | 1113 | /note= "potential phosphorylation site" |
| FT | Modified-site | 1113 | /note= "potential phosphorylation site" |
| XX | | | |
| PN | WO200006728-A2. | | |
| PD | 10-FEB-2000. | | |
| XX | | | |
| PF | 28-JUL-1999; | 99WO-US17132. | |
| XX | | | |
| PR | 28-JUL-1998; | 98US-0123494. | |
| PR | 14-SEP-1998; | 98US-0152814. | |
| PR | 14-OCT-1998; | 98US-0173482. | |
| PR | 03-NOV-1998; | 98US-0106889. | |
| PR | 19-NOV-1998; | 98US-0109093. | |
| PR | 22-DEC-1998; | 98US-0113796. | |
| PR | 12-JAN-1999; | 99US-0173482. | |
| PR | 12-JAN-1999; | 99US-0229005. | |
| XX | | | |
| PA | (INCY-) INCYTE PHARM INC. | | |
| XX | | | |
| PI | Hillman JL, Lal P, Tang YT, Corley NC, Guegler KU, Baughn MR, | | |
| PI | Patterson C, Bandman O, Au-Young J, Gorgone GA, Yue H, Azimzal Y, | | |
| PI | Reddy R, Lu DAM, Shih LT: | | |
| DR | N-PSDB: Z46153. | | |
| XX | WPI: 2000-183125/16. | | |
| XX | | | |
| XX | New human phosphorylation effectors useful for the diagnosis, treatment | | |
| PT | and prevention of proliferative, immune and neuronal disorders - | | |
| PT | | | |
| XX | Claim 1; Page 98-100; 142pp; English. | | |
| PS | | | |
| CC | Y66769-95 and Y68797-99 represent human phosphorylation effectors (PHSP), | | |
| CC | designated PHSP1-PHSP31 (the protein sequence for PHSP28 is not given | | |
| CC | in the specification). The sequences were isolated from cDNA libraries | | |
| CC | prepared from various human tissues. The PHSP proteins are useful for | | |
| CC | the diagnosis, treatment and prevention of proliferative disorders, | | |
| CC | immune disorders and neuronal disorders. The PHSP proteins form | | |
| CC | pharmaceutical compositions which useful for treating or preventing | | |
| CC | disorders associated with decreased PHSP expression/activity. PHSP | | |
| CC | antagonists are useful for treating or preventing disorders associated | | |
| CC | with increased PHSP expression/activity. | | |
| XX | | | |
| XX | Sequence 1135 AA: | | |
| SQ | | | |
| | Query Match | 10.9%; Score 143; DB 21; Length 1135; | |
| | Best Local Similarity 19.0%; Pred. No. 0.00068; | | |
| | Matches 66; Conservative 41; Mismatches 83; Indels 158; Gaps 9; | | |
| OY | 21 LDERRRRROE---EMEKVRKPEDDECEPEEYVDPRST----- | 54 | |
| Db | : : : : : : : : : : | | |
| Db | 306 idrttkrkgeketeyegysgeeeevpegepsaiynvgestlrrdflrlqgenke 365 | | |
| OY | 55 -----YERLOEQCDRKQOEYERDFPKKNMWRGIDEDDETNTLDEVSRQOELERQRRE 107 | | |
| Db | : : : : : : : : : : : : : : : : : : | | |
| Db | 366 rsealrrtqqllqeqqlrteqeykrgl-----laerqkrleqgkeqrtrlleeqqrter 417 | | |
| OY | 108 ELKELEKEYRNMLKKKGISOENKKVEYEKKLTVPRIETKMKFSQAKLLAGAVKHKSSESGNS 167 | | |
| Db | 418 earrqe-----regrrrrqeekrrlleelerrrrkeeeerrraaeekrrirvereqey 467 | | |
| OY | 168 VKRLKPD-----PEPDDKKNQEPSSCKSLGNTSLSGPSIHC 202 | | |
| Db | 468 lrrtqleeegrnhlevlqgqllqegamlldhxrphpbqhsqppp-----qgsrskpsfha 522 | | |
| OY | 203 P-----SAACIGIL--- 212 | | |
| Db | 523 pepkahyepadratevpyrttsrsplvsrrdsplqgsqsgnsqaqgrmstlslepriwer 582 | | |

[illegible]


```

Db 425 reqrrrgeekrrleerlrrekeerrrreerereqeyrrrqlleeqrhlvlg 484
QY 148 -----SQAKLAGAVKHKSSSGNSVKKRLKP-----PEPDDKNQEPSS 186
Db 485 qlleqgamll--ecrvremeehrqaelrqlqeqagayllslqdhrrpnpqhsqpppp 542
OY 187 CKSLGNTSLSGPSIHCP 203
Db 543 -----qgerskpsrthap 554

RESULT 6
Y55954
ID Y55954 standard: Protein; 1233 AA.
AC Y55954;
DT 18-FEB-2000 (first entry)
XX
DE Mouse STE20-related protein kinase NIK_m.
XX
KW Antithrombotic; antithrombotic; antiinflammatory; antiallergic; osteopathic;
KW antipsoriatic; antiarteriosclerotic; antiasthmatic; immunosuppressive;
KW neuroprotective; cardiant; cerebroprotective; cytoprotective; antidiabetic;
KW vulnery; STE20; protein kinase; STLK2; STLK3; STLK4; STLK5; STLK6; STLK7;
KW ZC1, ZC2, ZC4, KHS2, SUU1, SUU3, GEX2, PAK4, PAK5; antagonist;
KW antibody; gene therapy; rheumatoid arthritis; arteriosclerosis; asthma;
KW inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis;
KW rhinitis; autoimmunity; organ transplantation; multiple sclerosis;
KW myocardial infarction; cardiovascular disease; stroke; renal failure;
KW oxidative stress-related neurodegenerative disorder; Parkinson's disease;
KW amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy;
KW ischemic disorder; inflammation; diabetes mellitus; fibrosis; mitosis;
KW mesangial disorder; growth regulation; wound healing; T cell activation;
KW immunosuppressant.
XX
OS Mus sp.
XX
PN MO9953036-A2.
XX
PD 21-OCT-1999.
XX
PF 13-APR-1999; 99WO-US08150.
XX
PR 14-APR-1998; 98US-0081784.
XX
PA (SUGF-) SUGEN INC.
XX
PI Plowman G, Martinez R, Whyte D;
XX
DR WPI; 1999-611301/52.
XX
PT Novel kinase-related polypeptides used for the diagnosis and treatment
XX of kinase-related diseases and disorders
XX
PS Disclosure; Page 339-343; 387pp; English.
XX
XX
CC This sequence represents a novel STE20-related protein kinase. The
CC invention relates to nucleic acid molecule encoding a kinase polypeptide
CC selected from STLK2, STLK3, STLK4, STLK5, STLK6, STLK7, ZC1, ZC2, ZC3,
CC ZC4, KHS2, SUU1, SUU3, GEX2, PAK4 and PAK5. The proteins are used to
CC identify agonists and antagonists, and to raise antibodies. The
CC polynucleotides are useful in gene therapy protocols. The polynucleotides,
CC polypeptides, antibodies, antagonists and agonists may be used to treat
CC diseases such as immune-related disorders and diseases (e.g. rheumatoid
CC arthritis, arteriosclerosis, chronic inflammatory bowel disease (e.g.
CC Crohn's disease), asthma, osteoarthritis, psoriasis, atherosclerosis,
CC rhinitis, autoimmunity, and organ transplantation, chronic inflammatory
CC peptic disease, multiple sclerosis, organ transplantation, myocardial
CC infarction, cardiovascular disease, stroke, renal failure, oxidative
CC stress-related neurodegenerative disorders (e.g. amyotrophic lateral
CC sclerosis, Parkinson's disease and Leigh syndrome), cancer,

```

```

CC cardiomyopathies, ischemic disorders, inflammatory disorders, diabetes
CC mellitus, fibrotic and mesangial disorders. The proteins may also be
CC useful for cell growth regulation (e.g. in wound healing), T cell
CC activation, mitosis control, and as immunosuppressants.
XX
SO Sequence 1233 AA:

Query Match 10.6%; Score 139; DB 20; Length 1233;
Best Local Similarity 23.6%; Pred. No. 0.0017;
Matches 55; Conservative 37; Mismatches 75; Indels 66; Gaps 8;

OY 21 LDERKKRROE---EWKVRKPPDPEPCPEVYVDPRL----- 54
Db 305 idtrkrkgedeteysgeeeevpegepsivvpeastlrrflrlqgenke 364
OY 55 -----YERLQROKORQOYEEOQ-----KKRNVKGLDEDETFLDEVSRQOE 98
Db 365 rsealrrqqlleqqrlregeykrqlaerqrlleqqkeqrrrleeqqr-erearrqge 423
OY 99 LIEKORKEE---LKELEKTRNNLKKVGIQENKKEVEKLTIV--KPIETKKN---FSQA 150
Db 424 reqrrrgeekrrleerlrrekeerrrreerereqeyrrrqlleeqrhlvlg 483
OY 151 KLAGAVKHKSSSGNSVKKRLKPDPDDKNQEPSSCKSLGNTSLSGPSIHCP 203
Db 484 qlleqgamllhthrnpqppppqqr-----skpsrthap 521

RESULT 7
Y58500
ID Y58500 standard: Protein; 1162 AA.
AC Y58500;
DT 10-APR-2000 (first entry)
XX
DE HHV8 ORF 73 protein, SEQ ID NO:21.
XX
KW HHV8; detection; diagnosis; Kaposi's sarcoma; AIDS; immunogen;
KW antigen.
XX
OS Human herpesvirus type 8.
XX
FH Key Location/Qualifiers
FT Misc-difference 96
FT /Label= unknown
XX
PN MO9961909-A2.
XX
PD 02-DEC-1999.
XX
PF 26-MAY-1999; 99WO-US11407.
XX
PR 26-MAY-1998; 98US-0086695.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pau CP;
XX
DR WPI; 2000-097142/08.
XX
PT New methods and compositions for the detection of human herpesvirus
XX
PS Claim 2; Page 59-62; 68pp; English.
XX
XX
CC Sequences Y58480-Y58532 represent immunogenic polypeptides derived from
CC human herpes virus type 8 (HHV8, a gammaherpesvirus). HHV8 plays an
CC important role in the pathogenesis of AIDS-related Kaposi's sarcoma. The
CC invention relates to a novel method of detecting the presence of human
CC herpesvirus 8 in a biological sample using peptides representative of
CC dominant antigenic regions of HHV8. The method comprises contacting one
CC or more isolated, immunogenic HHV8 peptides with an antibody-containing

```



```

QY 124 ISOENK-----KEYEKRTVTPRIETKNFSGAKLLAGAVKHKSSSEGSVYKRLK 112
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 1011 heeekvtheeenvtyeeeek-vmkkkk1mkvkkkkkkkkkkggamm--- 106

```

```

QY 124 ISOENK-----KEYEKKTIVPITFKNFSQAKLLAAGVHKHSSSENSVYRLK 172
Db 1011 heeeekvttheeenvtlyeeeeeekytheeeek-vmkkkkkmxykkkkkkkkkggamm--- 106
QY 173 PDPEPDDKNOEP--SSCKSLGNTSLSPSHCES 204
Db 1067 --vpgsngsdplectcrg-----awhwps 1088

```

[illegible]

| | | | |
|--------|-------------|---|------|
| QY | 124 | ISQCNK-----KEVEKLTVPPIETKKNFSQAKLLAAGVKKHSSSEGSNVRLK | 172 |
| | | : : : : : : : : : : : | |
| Db | 1011 | heeeekvtheeeenvlyeeneekvtheeeek-vmkkkkImkykkkkkkkkkgagam--- | 106 |
| QY | 173 | PDPEPDDKNOEP--SSCKSLGNTSLSGPSIHCS | 204 |
| | | : : : : : : : : : : | |
| Db | 1067 | ---vpgsasnqplestctrg-----awhwps | 1088 |
| RESULT | 11 | | |
| Y85263 | | | |
| ID | Y85263 | standard; Protein, 1360 AA. | |
| XX | | | |
| AC | Y85263; | | |
| XX | | | |
| OT | 29-JUN-2000 | (first entry) | |

| | | | |
|----|------|---|------|
| QY | 124 | ISOQNK-----KEVEKATLTPVPIETKNKFSGAKLLAAVKHKKSSSEGSNSVRLK | 172 |
| Db | 1011 | heeeekvttheeekvtyeeeeeekvttheeek-vmkkkktimkykkkkkkkkkggamm--- | 106 |
| QY | 173 | PDPPEDDKNGEP--SSCKSLGNTSLSGPSIHCS | 204 |
| Db | 1067 | --vpgsnsdplstctcrg-----awhwps | 1088 |

| | |
|--------|---|
| RESULT | 11 |
| Y85263 | |
| ID | Y85263 standard; Protein; 1360 AA. |
| XX | |
| AC | Y85263; |
| XX | |
| DJ | 29-JUN-2000 (first entry) |
| XX | |
| DE | Human protein kinase KIAA0551 amino acid sequence. |
| XX | |
| KX | Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation; |
| KW | chronic pain; neurodegenerative disorder; neurotraumatic disorder; |

| | | | |
|--------|---|---|------|
| QY | 124 | ISQGNK-----KEVEKATLVPIETKKNFSQAKLIAAVKHKSSGSGNSVRLK | 172 |
| | | : : : : : : : : : : : : : : | |
| Dbl | 1011 | heeekvttheeekvtyeeeeekvttheeek-vmkkkkikmkkkkkkkkkkggam--- | 106 |
| QY | 173 | PDPEPDDKNOEP--SSCKSLGNTSLGPSIHCP5 | 204 |
| | | : : : : : : : : : : : : : : | |
| Dbl | 1067 | --Vpgssnsdqllestcrg-----awhwps | 1088 |
| RESULT | 11 | | |
| Y85263 | | | |
| ID | Y85263 | standard; Protein; 1360 AA. | |
| XX | | | |
| XX | Y85263; | | |
| AC | | | |
| XX | | | |
| DT | 29-JUN-2000 | (first entry) | |
| XX | | | |
| DE | Human protein kinase KIAA0551 | amino acid sequence. | |
| XX | | | |
| KW | Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation; | | |
| KW | chronic pain; neurodegenerative disorder; neurotraumatic disorder; | | |
| KW | Parkinson's disease; Alzheimer's disease; ischaemic disease. | | |
| XX | | | |
| OS | Homo sapiens. | | |
| XX | | | |

| | | | |
|--------|---|--|------|
| QY | 124 | ISOENK-----KEVEKKFTVVPDIFTKNFSQAKLIAAGVKHKSSGNSVRLK | 172 |
| Dd | 1011 | heeekvttheeekvtyeeeeeekvttheeek-vmkkkkikmykkkkkkkggam--- | 106 |
| OY | 173 | PDPPEDDKNDEP--SSCKSLGNTSLSGPSIHCP5 | 204 |
| Dd | 1067 | --vpgsnsdqplestcrg-----awhmws | 1088 |
| RESULT | 11 | | |
| Y85263 | | | |
| ID | Y85263 | standard; Protein; 1360 AA. | |
| XX | | | |
| AC | Y85263; | | |
| XX | | | |
| DT | 29-JUN-2000 | (first entry) | |
| XX | | | |
| DE | Human protein kinase KIAA0551 amino acid sequence. | | |
| XX | | | |
| KW | Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation; chronic pain; neurodegenerative disorder; neurotraumatic disorder; | | |
| KM | Parkinson's disease; Alzheimer's disease; ischaemic disease. | | |
| XX | | | |
| OS | Homo sapiens. | | |
| XX | | | |
| FH | Key | Location/Qualifiers | |
| FT | Misc-difference 290 | | |
| XX | /note= "Encoded by GA" | | |
| XX | WO200015805-A1. | | |
| XN | | | |

| QY | 124 | ISOENK-----KEVEKLTVPDFTKKNFSQAKLLAGAVKHKSSSESGNSVRLK | 172 |
|--------|---|---|------|
| Dd | 1011 | heeeekvttheeeenvtyeeeeeekvttheeeek-vmkkkkkimykkkkkkkkkkggagm--- | 106 |
| QY | 173 | PDPPEDDKNOEP--SSCKSLGNTSLSPSHCS | 204 |
| Dd | 1067 | --vpgsnsqblestcrq-----awhmps | 1088 |
| RESULT | 11 | | |
| ID | Y85263 | | |
| XX | Y85263 | standard; Protein; 1360 AA. | |
| XX | Y85263 | | |
| XX | 29-JUN-2000 | (first entry) | |
| DE | Human protein kinase KIAA0551 | amino acid sequence. | |
| XX | XX | | |
| KW | Protein kinase: KIAA0551; neuropathy; neuropathic pain; inflammation; | | |
| KW | chronic pain; neurodegenerative disorder; neurotraumatic disorder; | | |
| KW | Parkinson's disease; Alzheimer's disease; ischaemic disease. | | |
| XX | XX | | |
| OS | Homo sapiens. | | |
| XX | XX | | |
| XX | Key | Location/Qualifiers | |
| FT | Misc-difference | 290 | |
| FT | /note= | "Encoded by GA" | |
| PN | W0200015805-A1. | | |
| XX | XX | | |
| DD | 23-MAR-2000 | | |

| | | | |
|---------------|---------------------|---|------|
| QY | 124 | ISOENK-----KEVEKKTVNPITFKNKSQAOKLLAAGVKKHSSSEGNVRLK | 172 |
| | | : : : : : : : : : : : : : : | |
| Db | 1011 | heeeekvtheeeenvtyeeneekvttheeeek-vmkkttkimykktkkkkkkggam--- | 106 |
| OY | 173 | PDEPPDDKNOEP--SCCKSLGNTSLSGPSTHCPS | 204 |
| | | : : : : : : : : : : : : : : | |
| Dd | 1067 | --vpgsnsdqplestctgq-----awhwps | 1088 |
| RESULT 11 | | | |
| ID | Y85263 | standard; Protein, 1360 AA. | |
| XX | AC | Y85263; | |
| XX | DT | 29-JUN-2000 (first entry) | |
| XX | DE | Human protein kinase KIAA0551 amino acid sequence. | |
| KW | KM | Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation; chronic pain; neurodegenerative disorder; neurotraumatic disorder; | |
| XX | KW | Parkinson's disease; Alzheimer's disease; ischaemic disease. | |
| XX | OS | Homo sapiens. | |
| XX | FH | Key location/Qualifiers | |
| FT | Misc-difference 290 | /note= "Encoded by GA" | |
| XX | PN | WO200015805-A1. | |
| XX | PD | 23-MAR-2000. | |
| XX | PF | 10-SEP-1999; 99MO-GB03017. | |
| XX | PR | 10-SEP-1998; 98GB-0019779. | |

| QY | 124 | ISOCONK-----KEVEKKTFTVPDIFTKNFSQAKLLAAGVKKHSSGSGNSVRLK | 172 |
|--------|---------------------------------|---|------|
| Db | 1011 | heeeekvttheeekvtyteeekkvtheeek-vmkkkkikmkkkkkkkkggam--- | 106 |
| QY | 173 | PDPEPDDKNOEP--SSCKSLGNTSLSGPSIHCP5 | 204 |
| Db | 1067 | --vpgsnsdqplestcrg-----awhmps | 1088 |
| RESULT | 11 | | |
| Y85263 | | | |
| ID | Y85263 | standard; Protein; 1360 AA. | |
| XX | | | |
| AC | Y85263; | | |
| XX | | | |
| DT | 29-JUN-2000 | (first entry) | |
| XX | | | |
| DE | | Human protein kinase KIAA0551 amino acid sequence. | |
| XX | | | |
| KW | | Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation; | |
| KM | | chronic pain; neurodegenerative disorder; neurotraumatic disorder; | |
| KW | | Parkinson's disease; Alzheimer's disease; ischaemic disease. | |
| XX | | | |
| OS | | Homo sapiens. | |
| XX | | | |
| FH | Key | Location/Qualifiers | |
| FT | Misc-difference 290 | | |
| FT | /note="Encoded by GA" | | |
| N | WO200015805-A1. | | |
| XX | | | |
| PD | 23-MAR-2000. | | |
| XX | | | |
| PF | 10-SEP-1999; | 99MO-GB03017. | |
| XX | | | |
| PR | 10-SEP-1998; | 98GB-0019779. | |
| XX | 29-MAR-1999; | 99GB-0007261. | |
| PA | (SMIK) SMITHKLINE BEECHAM PLC. | | |

```

QY      124  ISOCON-----KEVEKTLTVPIFTKNKFSGAKLLAGAVKKHSSSEGSNVRLK 172
           :|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1011 heeeekvtheeenvtyeeeeekvtheeeeek-vmkxkkkimykrkkkkkkkggamm--- 106
           |   |   |   |   |   |   |   |   |   |   |   |   |   |
QY      173 PDPEPDDKNGEP--SCKSLGANTSLSGPSIHCP8 204
           |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      1067 ---vpgsnsdplstctrg-----awhmps 1088
           |   |   |   |   |   |   |   |   |   |   |   |   |   |

RESULT  11
Y85263
AC      Y85263 standard; Protein; 1360 AA.
XX
XX      Y85263;
XX
DT      29-JUN-2000 (first entry)
XX
DE      Human protein kinase KIAA0551 amino acid sequence.
XX
KW      Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation;
KM      chronic pain; neurodegenerative disorder; neurotransmitter disorder;
KM      Parkinson's disease; Alzheimer's disease; ischaemic disease.
XX
OS      Homo sapiens.
XX
FH      Key Location/Qualifiers
FT      Misc-difference 290
FT      /note= "Encoded by GA"
XX
PN      WO200015805-A1.
XX
XX      23-MAR-2000.
XX
XX      PD
XX      PF      10-SEP-1999; 99WO-GB3017.
XX
XX      PR      10-SEP-1998; 98GB-0019779.
XX      PR      29-MAR-1999; 99GB-0007261.
XX
XX      PA      (SMIK ) SMITHKLINE BEECHAM PLC.
XX
XX      PI      Bingham S, Case P, Lawson SN, Newton RA, Rausch OL, Relth AD;
XX      PI      Sanger GU;
XX
XX      WPI: 2000-271443/23.
XX
XX      N-PSDB: A10669.

```

| QY | 124 | ISOCHK | ----- | KEVKKLTFTVPIPTKKNFSGAKLLAGAVKHKSSSEGNVRLK | 172 |
|--------|-------------|--|--|---|-----|
| QY | 1011 | heeeekvtheeeekvtheeeekvtheeeek | -vmkkttkimkykkkkkkkkggagm--- | 106 | |
| QY | 173 | PDPEPDKNOEP-- | SSCKSLGNTSLSGPSIHCP5 | 204 | |
| Db | 1067 | --vpssnsdplestcrq | -----awhwps | 1088 | |
| RESULT | 11 | | | | |
| ID | Y85263 | | | | |
| XX | Y85263 | standard; Protein; 1360 | AA. | | |
| XX | Y85263; | | | | |
| XX | 29-JUN-2000 | (first entry) | | | |
| XX | | | | | |
| XX | | Human protein kinase KIAA0551 amino acid sequence. | | | |
| KW | | Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation; | | | |
| KW | | chronic pain; neurodegenerative disorder; neurotraumatic disorder; | | | |
| KW | | Parkinson's disease; Alzheimer's disease; ischemic disease. | | | |
| XX | | | | | |
| OS | | Homo sapiens. | | | |
| XX | | | | | |
| FH | | Key | Location/Qualifiers | | |
| FT | | Misc-difference | 290 | | |
| FT | | /note= | "Encoded by GA" | | |
| XX | | | | | |
| PN | | WO200015805-A1. | | | |
| XX | | | | | |
| PD | | 23-MAR-2000. | | | |
| XX | | | | | |
| PF | | 10-SEP-1999; | 99WO-GB03017. | | |
| XX | | | | | |
| PR | | 10-SEP-1998; | 98GB-0019779. | | |
| PR | | 29-MAR-1999; | 99GB-0007261. | | |
| XX | | | | | |
| PA | | (SMIR) | SMITHKLINE BEECHAM PLC. | | |
| PI | | Bingham S, | Case P, Lawson SN, Newton RA, Rausch OL, Relth AD; | | |
| PI | | Sanger GJ; | | | |
| DR | | WPI; 2000-271443/23. | | | |
| XX | | N-PSDB; A10669. | | | |
| PT | | Isolated human KIAA0551 polynucleotide and polypeptide, useful for | | | |
| PT | | treating e.g. neuropathies, neuropathic pain, inflammatory and chronic | | | |
| PT | | pain and neurodegenerative conditions - | | | |

| QY | 124 | ISOCON | -----KEVEKKTIVVPITFKNF | SOAKLLAAGAVKHKSSSEGN | SVRLK | 172 |
|----|--|---|--------------------------|----------------------|-------|-----|
| Db | 1011 | heeeekvtheeeenvtyeeeeeekvtheeeek | -vmkkttkimykktkkkkkkggam | --- | 106 | |
| QY | 173 | PDPEPDDKNOEP | --SCKSLGNTSLSGPS | IHCPS | 204 | |
| Db | 1067 | --VPgsnsdplestcrq | -----awhps | 1088 | | |
| | | RESULT | 11 | | | |
| ID | Y85263 | standard; Protein; 1360 | AA. | | | |
| XX | Y85263; | | | | | |
| AC | Y85263; | | | | | |
| XX | 29-JUN-2000 | (first entry) | | | | |
| DT | 29-JUN-2000 | (first entry) | | | | |
| XX | | | | | | |
| DE | | Human protein kinase KIAA0551 amino acid sequence. | | | | |
| XX | | | | | | |
| KW | | Protein kinase: KIAA0551; neuropathy; neuropathic pain; inflammation; | | | | |
| KW | | chronic pain; neurodegenerative disorder; neurotraumatic disorder; | | | | |
| KW | | Parkinson's disease; Alzheimer's disease; Ischaemic disease. | | | | |
| XX | | | | | | |
| OS | | Homo sapiens. | | | | |
| XX | | | | | | |
| FH | Key | Location/Qualifiers | | | | |
| FT | Misc-difference 290 | /note= "Encoded by GA" | | | | |
| XX | | | | | | |
| PN | WO200015805-A1. | | | | | |
| XX | | | | | | |
| PD | 23-MAR-2000. | | | | | |
| XX | | | | | | |
| PE | 10-SEP-1999; | 99WO-GB03017. | | | | |
| XX | | | | | | |
| PR | 10-SEP-1998; | 98GB-0019779. | | | | |
| PR | 29-MAR-1999; | 99GB-0007261. | | | | |
| XX | | | | | | |
| PA | (SMIK) SMITHKLINE BEECHAM PLC. | | | | | |
| PI | Bingham S, | Case P, Lawson SN, Newton RA, Rausch OL, Relth AD; | | | | |
| PI | Sanger GJ; | | | | | |
| DR | WPI: 2000-271443/23. | | | | | |
| DR | N-PSDB; A10669. | | | | | |
| XX | | | | | | |
| PT | Isolated human KIAA0551 polynucleotide and polypeptide, useful for | | | | | |
| PT | treating e.g. neuropathies, neuropathic pain, inflammatory and chronic | | | | | |
| PT | pain and neurodegenerative conditions - | | | | | |
| DS | Claim 2; Page 41; 48pp; English. | | | | | |

| | | | | |
|--------|--|-----------------------------------|--|------|
| QY | 124 | ISOCON | -----KEVEKKTIVPIFTKKNFSQAKLLAAGVKKHSSSEGSNVRLK | 172 |
| Db | 1011 | heeeekytheeenvtyeeeeeekrvtheeeeek | -vmkkkkkilmkykkkkkkkkkggam | 106 |
| QY | 173 | PDPEPDKRQEP | --SCKSLGANTSLSGPSIHCP | 204 |
| Db | 1067 | --vpgssnsdplstcrtg | -----awhmps | 1088 |
| RESULT | 11 | | | |
| ID | Y85263 | | | |
| XX | Y85263 | standard; Protein; 1360 AA. | | |
| XX | Y85263; | | | |
| DT | 29-JUN-2000 | (first entry) | | |
| DE | Human protein kinase KIAA0551 amino acid sequence. | | | |
| XX | | | | |
| KW | Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation; | | | |
| KW | chronic pain; neurodegenerative disorder; neurotraumatic disorder; | | | |
| KW | Parkinson's disease; Alzheimer's disease; ischaemic disease. | | | |
| XX | | | | |
| OS | Homo sapiens. | | | |
| XX | | | | |
| FH | Key | Location/Qualifiers | | |
| FT | Misc-difference 290 | | | |
| FT | /note= "Encoded by GA" | | | |
| XX | WP200015805-A1. | | | |
| XX | 23-MAR-2000. | | | |
| XX | 10-SEP-1999; | 99MO-GB03017. | | |
| XX | 10-SEP-1998; | 98GB-0019779. | | |
| XX | 29-MAR-1999; | 99GB-0007261. | | |
| XX | (SMIK) SMITHKLINE BEECHAM PLC. | | | |
| XX | Bingham S, Case P, Lawson SN, Newton RA, Rausch OL, Reith AD; | | | |
| XX | Sanger GJ; | | | |
| XX | WPI; 2000-271443/23. | | | |
| XX | N-PSDB; A10669. | | | |
| XX | | | | |
| PT | Isolated human KIAA0551 polynucleotide and polypeptide, useful for | | | |
| PT | treating e.g. neuropathies, neuropathic pain, inflammatory and chronic | | | |
| PT | pain and neurodegenerative conditions - | | | |
| XX | | | | |
| PS | Claim 2; Page 41; 48pp; English. | | | |
| CC | This sequence represents the human protein kinase KIAA0551 amino acid | | | |
| CC | sequence. The nucleotide sequence was isolated from a human foetal brain | | | |
| CC | cDNA library. The rat KIAA0551 mRNA is upregulated in dorsal root gang | | | |
| CC | during sciatic neuropathy (a procedure accompanied by increased | | | |
| CC | sensitivity to somatic pain) which indicates a role for KIAA0551 in the | | | |

| | | | |
|--------|--|---|------|
| QY | 124 | ISOENK-----KEVEKKTIVPIETKKNFSQAKLLAAVKKHSSSESGNSVRLK | 172 |
| Db | 1011 | heeeekvtheeeenvtyeeeeekvtheeeek-vmkkkkkilmkykkkkkkkkkggamm--- | 106 |
| QY | 173 | PDPEPDKNDPE--SSCKSLGNTSLSGPSIHCS | 204 |
| Db | 1067 | --vpssnsdplstcrg-----awhps | 1088 |
| RESULT | 11 | | |
| ID | Y85263 | | |
| AC | Y85263 | standard; Protein; 1360 AA. | |
| XX | Y85263; | | |
| DT | 29-JUN-2000 | (first entry) | |
| DE | Human protein kinase KIAA0551 amino acid sequence. | | |
| XX | | | |
| KW | Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation; | | |
| KW | chronic pain; neurodegenerative disorder; neurotraumatic disorder; | | |
| KW | Parkinson's disease; Alzheimer's disease; ischaemic disease. | | |
| XX | | | |
| OS | Homo sapiens. | | |
| XX | | | |
| FH | Key | Location/Qualifiers | |
| FT | Misc-difference 290 | | |
| FT | /note= "Encoded by GA" | | |
| XX | W0200015805-A1. | | |
| XX | 23-MAR-2000. | | |
| XX | 10-SEP-1999; | 99WO-GB03017. | |
| XX | 10-SEP-1998; | 98GB-0019779. | |
| PR | 29-MAR-1999; | 99GB-0007261. | |
| XX | | | |
| FA | (SMIK) SMITHKLINE BEECHAM PLC. | | |
| XX | | | |
| PI | Bingham S, Case P, Lawson SN, Newton RA, Rausch OL, Reith AD; | | |
| PI | Sanger GJ; | | |
| DR | WPI: 2000-271443/23. | | |
| XX | N-PSDB; A10669. | | |
| PT | Isolated human KIAA0551 polynucleotide and polypeptide, useful for | | |
| PT | treating e.g. neuropathies, neuropathic pain, inflammatory and chronic | | |
| PT | pain and neurodegenerative conditions - | | |
| XX | | | |
| PS | Claim 2; Page 41; 48pp: English. | | |
| XX | | | |
| CC | This sequence represents the human protein kinase KIAA0551 amino acid | | |
| CC | sequence. The nucleotide sequence was isolated from a human foetal brain | | |
| CC | cDNA library. The rat KIAA0551 mRNA is upregulated in dorsal root gan | | |
| CC | during sciatic neuropathy (a procedure accompanied by increased | | |
| CC | sensitivity to somatic pain) which indicates a role for KIAA0551 in t | | |
| CC | regulation of molecular processes associated with neuropathy and | | |
| CC | neuropathic pain. The KIAA0551 protein may be used for treating | | |

| | | | |
|--------|--------|--|------|
| QY | 124 | ISOCON-----KEVEKTKLVTPITFKNFSQAKLAAVKKHSSSESNVRLK | 172 |
| Dy | 1011 | heeekvttheeenvtyeeekvtheeek-vmkkkkilmkykkkkkkkggam--- | 106 |
| QY | 173 | PDEPPDDKNQEP--SSCKSLGNTSLSGPSIHCP5 | 204 |
| Dy | 1067 | --vpgsnsdplestcrq-----awhwps | 1088 |
| RESULT | 11 | | |
| ID | Y85263 | Y85263 standard; Protein; 1360 AA. | |
| XX | AC | Y85263; | |
| XX | DT | 29-JUN-2000 (first entry) | |
| XX | DE | Human protein kinase KIAA0551 amino acid sequence. | |
| KW | KW | Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation; | |
| KM | KM | chronic pain; neurodegenerative disorder; neurotraumatic disorder; | |
| KW | KW | Parkinson's disease; Alzheimer's disease; Ischaemic disease. | |
| OS | OS | Homo sapiens. | |
| XX | XX | | |
| FH | FH | Key Location/Qualifiers | |
| FT | FT | Misc-difference 290 | |
| FT | FT | /note= "Encoded by GA" | |
| PN | PN | WO200015805-A1. | |
| PD | PD | 23-MAR-2000. | |
| XX | XX | | |
| PF | PF | 10-SEP-1999; 99MO-GB03017. | |
| XX | XX | | |
| PR | PR | 10-SEP-1998; 98GB-0019779. | |
| PR | PR | 23-MAR-1999; 99GB-0007261. | |
| XX | XX | (SMIK) SMITHKLINE BEECHAM PLC. | |
| PA | PA | Bingham S, Case P, Lawson SN, Newton RA, Rausch OL, Relch AD; | |
| PI | PI | Sanger GJ; | |
| DR | DR | WPI; 2000-271443/23. | |
| XX | XX | N-PDB; A10669. | |
| PT | PT | Isolated human KIAA0551 polynucleotide and polypeptide, useful for | |
| PT | PT | treating e.g. neuropathies, neuropathic pain, inflammatory and chroni- | |
| XX | XX | c pain and neurodegenerative conditions - | |
| PS | PS | Claim 2; Page 41; 48pp; English. | |
| XX | XX | This sequence represents the human protein kinase KIAA0551 amino acid | |
| CC | CC | sequence. The nucleotide sequence was isolated from a human foetal br | |
| CC | CC | CNA library. The rat KIAA0551 mRNA is upregulated in dorsal root gan | |
| CC | CC | during sciatic neuropathy (a procedure accompanied by increased | |
| CC | CC | sensitivity to somatic pain) which indicates a role for KIAA0551 in t | |
| CC | CC | regulation of molecular processes associated with neuropathy and | |
| CC | CC | neuropathic pain. The KIAA0551 protein may be used for treating | |
| CC | CC | neuropathies, neuropathic pain, inflammatory and chronic pain, | |
| CC | CC | neurodegenerative conditions such as Parkinson's disease and Alzheim | |
| CC | CC | disease, and neurotraumatic disease or Ischaemic disease damage in | |
| XX | XX | cardiac tissue. | |

| Query Match | 9.6% | Score 126 | DB 21 | Length 1360 |
|---|---|-----------|-------|-------------|
| QY 124 | ISOCONT-----KEVEKTLTVPIFTKNGFSQAKLLAAVKKSSSESNVRLK | 172 | | |
| Db 1011 | heeeekvttheeenvtyeeeeekvttheeek-vmkkkkimkykkkkkkkkkggamm--- | 106 | | |
| QY 173 | PDPEDDKNGEP--SCKSLGNTSLSGPSIHCS | 204 | | |
| Db 1067 | --vpgsnsdplestcrg-----awhmps | 1088 | | |
| RESULT 11 | | | | |
| ID Y85263 | | | | |
| XX Y85263 standard; Protein; 1360 AA. | | | | |
| XX AC Y85263; | | | | |
| XX 29-JUN-2000 (first entry) | | | | |
| XX DE Human protein kinase KIAA0551 amino acid sequence. | | | | |
| XX KW Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation; | | | | |
| XX KW chronic pain; neurodegenerative disorder; neurotraumatic disorder; | | | | |
| XX KW Parkinson's disease; Alzheimer's disease; ischaemic disease. | | | | |
| XX OS Homo sapiens. | | | | |
| XX FH Key Location/Qualifiers | | | | |
| XX FT Misc-difference 290 | | | | |
| XX FT /note= "Encoded by GA" | | | | |
| XX WO200015805-A1. | | | | |
| XX PD 23-MAR-2000. | | | | |
| XX XX 10-SEP-1999; 99WO-GB03017. | | | | |
| XX XX 10-SEP-1998; 98GB-0019779. | | | | |
| XX PR 29-MAR-1999; 99GB-0007261. | | | | |
| XX XX (SMIK) SMITHKLINE BEECHAM PLC. | | | | |
| XX PA Bingham S, Case P, Lawson SN, Newton RA, Rausch OL, Relth AD; | | | | |
| XX PI Sanger GJ; | | | | |
| XX DR N-PSDB; A10669. | | | | |
| XX WPI: 2000-271443/23. | | | | |
| XX PT Isolated human KIAA0551 polynucleotide and polypeptide, useful for | | | | |
| XX PT treating e.g. neuropathies, neuropathic pain, inflammatory and chronic | | | | |
| XX PT pain and neurodegenerative conditions - | | | | |
| XX XX | | | | |
| XX Claim 2; Page 41; 48pp; English. | | | | |
| XX XX This sequence represents the human protein kinase KIAA0551 amino acid | | | | |
| XX CC sequence. The nucleotide sequence was isolated from a human foetal brain | | | | |
| XX CC cDNA library. The rat KIAA0551 mRNA is upregulated in dorsal root ganglia | | | | |
| XX CC during sciatic neuropathy (a procedure accompanied by increased | | | | |
| XX CC sensitivity to somatic pain) which indicates a role for KIAA0551 in the | | | | |
| XX CC regulation of molecular processes associated with neuropathy and | | | | |
| XX CC neuropathic pain. The KIAA0551 protein may be used for treating | | | | |
| XX CC neuropathies, neuropathic pain, inflammatory and chronic pain, | | | | |
| XX CC neurodegenerative conditions such as Parkinson's disease and Alzheimer's | | | | |
| XX CC disease, and neurotraumatic disease or ischaemic disease damage in | | | | |
| XX CC cardiac tissue. | | | | |
| XX SQ Sequence 1360 AA; | | | | |

| | | |
|--|---------------------|---|
| QY | 124 | ISOCON-----KEVEKLTVPITFKNNFSQAALLAAVHKHSSSEGNYSVRLEK 172 |
| Dd | 1011 | heeeekrtheeenvtyeeeeeekvrtheeek-vmkkkkklmkykkkkkkkggam---106 |
| OY | 173 | PDEPPDKNQEP--SCKSLGANTSLSGPSIHCP8 204 |
| Dd | 1067 | --vpgssnsdplesctrg-----awhwps 1088 |
| RESULT 11 | | |
| ID | Y85263 | standard; Protein, 1360 AA. |
| XX | AC | Y85263; |
| XX | DT | 29-JUN-2000 (first entry) |
| XX | DE | Human protein kinase KIAA0551 amino acid sequence. |
| XX | KW | Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation; |
| XX | KW | chronic pain; neurodegenerative disorder; neurotraumatic disorder; |
| XX | KW | Parkinson's disease; Alzheimer's disease; ischaemic disease. |
| XX | OS | Homo sapiens. |
| XX | FH | Key Location/Qualifiers |
| FT | Misc-difference 290 | /note= "Encoded by GA" |
| XX | PN | WO200015805-A1. |
| XX | PD | 23-MAR-2000. |
| XX | PF | 10-SEP-1999; 99MO-GB03017. |
| XX | PR | 10-SEP-1998; 98GB-0019779. |
| XX | PR | 29-MAR-1999; 99GB-0007261. |
| XX | PA | (SMIK) SMITHKLINE BEECHAM PLC. |
| PI | Bingham S, | Case P, Lawson SN, Newton RA, Rausch OL, Reith AD; |
| PI | Sanger GJ; | |
| DR | N-PSDB: | A10669. |
| XX | PT | Isolated human KIAA0551 polynucleotide and polypeptide, useful for |
| XX | PT | treating e.g. neuropathies, neuropathic pain, inflammatory and chronic |
| XX | PT | pain and neurodegenerative conditions - |
| XX | PS | Claim 2; Page 41; 48pp; English. |
| XX | CC | This sequence represents the human protein kinase KIAA0551 amino acid |
| XX | CC | sequence. The nucleotide sequence was isolated from a human foetal brain |
| XX | CC | cDNA library. The rat KIAA0551 mRNA is upregulated in dorsal root ganglia |
| XX | CC | during sciatic neuropathy (a procedure accompanied by increased |
| XX | CC | sensitivity to somatic pain) which indicates a role for KIAA0551 in the |
| XX | CC | regulation of molecular processes associated with neuropathy and |
| XX | CC | neuropathic pain. The KIAA0551 protein may be used for treating |
| XX | CC | neuropathies, neuropathic pain, inflammatory and chronic pain, |
| XX | CC | neurodegenerative conditions such as Parkinson's disease and Alzheimer's |
| XX | CC | disease, and neurotraumatic disease or ischaemic disease damage in |
| XX | CC | cardiac tissue. |
| SO | Sequence | 1360 AA. |
| Query Match 9.6%; Score 126; DB 21; Length 1360; Best Local Similarity 23.1%; Pred. No. 0.025; Matches 54; Conservative 50; Mismatches 68; Indels 62; Gaps | | |
| Yy | 21 | LDERRRKRROEMEKVKRPDPPECEPEY-----YDPRSILYEKLQECKDKRQDEYEOR 74 |


```

OY      70 EQGFEEKNNVRGLDEDETFTNLEDEVSR-OQELIEKORREBELR-ETKEEYNINLKTKVGISOE 127
       ||| |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db     445 eellIar-----eelsrlqgetcqlseesvsyKaqllepIQhnlyqe---sqQ 487

OY      128 NKEVEKKTLVKPIPTFKNKFS-----QAKLGAAYKH-----KSSEGN----- 166
          ::| :|:|| | |   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db    468 elsmgmrlcmldletdmngmswsspgsvlryngatdyclsjsttssetanfnheagqm 547

OY      167 -----SVKRRLKPDPEDDKNOEPSSCKSLGNTSLS 196
           -|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-:-
Db    548 lesepthagessvrspselapsdvtcdeseavtvagnekvl 586


RESULT  14
W944A06 W944A06 standard; Protein; 897 AA.

XX AC W944A06;
XX DT 19-APR-1999 (first entry)
XX DE Murine eps15 protein.
XX KM Murine; eps15; epidermal growth factor receptor; EGFR; triple helix;
XX tyrosine kinase receptor; mitogenic signal transduction; detection;
XX malignant tissue.
XX OS Mus sp.
XX PN US5872219-A.
XX PD 16-FEB-1999.
XX PF 07-JUN-1995; 95US-0477389.
XX PR 22-JUL-1993; 93US-0095737.
XX PR 25-AUG-1992; 92US-093531L.
XX PR 07-JUN-1995; 95US-0477389.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI DiFiore PP, Fazioli F;
XX DR WPJ; 1999-166718/14.
XX DR N-PADB; X04192.
XX PT New anti-eps15 antibodies - used for detection of eps15, tyrosine
XX PT kinase receptor kinase activity and altered mitogenic signal
XX PT transduction
XX PS Claim 2; Column 37-42; 26pp; English.
XX CC The present invention describes antibodies which specifically bind to
CC human and murine eps (epidermal growth factor receptor pathway
CC substrate) 15. Also described are purified antibodies that specifically
CC bind to eps15 serving as a substrate for tyrosine phosphorylation
CC following epidermal growth factor receptor (EGFR) activation, where
CC the amino acids of eps15 hybridise under low stringency conditions to
CC the protein-encoding domain of human or murine eps15 polynucleotides.
CC The antibodies can be used to assay eps15 in samples. They can also be
CC used to determine tyrosine kinase receptor (TKR) activity in samples
CC and to detect altered mitogenic signal transduction, particularly in
CC malignant tissues. The present sequence represents murine eps15.
XX SC Sequence 897 AA;
```

| | | | |
|--------|--------|--|-----|
| Db | 390 | qkqyvgelqgldedqkqleeqdqyevrk-----kceeeqjlsslkaeltsgesqjssye | 444 |
| Oy | 70 | EEQEFKNMNVAGLDEDETNFLDEYSR-QQELIEKQREBEELK-ELKEYRNLIKVKYISOE | 127 |
| Db | 445 | eellkar-----eelsrlqgetaqleesvesvkaqllepqlqhqe---sqg | 487 |
| Oy | 128 | NKKEVEKKLVKPLETKNKS-----QAKLLAGAVKH-----KSESGN----- | 166 |
| Db | 488 | eissmqmrlemkdletdnngsnwsspsqyvlvngadycslsstssetanfhnehaegqn | 547 |
| Oy | 167 | -----SVKRLKPDPEDDKNQOEPSCKSLSGNTLS | 196 |
| Db | 548 | lseptngessvrspselapsdvtdeaseavtvagnkevt | 586 |
| RESULT | 15 | | |
| Y55932 | | | |
| ID | Y55932 | standard; Protein; 1297 AA. | |
| XX | AC | Y55932; | |
| XX | DT | 18-FEB-2000 (first entry) | |
| XX | DE | Human ZC2 protein. | |
| XX | XX | Antirheumatic; antiarthritic; antiinflammatory; antiallergic; osteoporotic; antiporiatic; antiarteriosclerotic; antiasthmatic; immunosuppressive; neuroprotective; cardiac; cerebroprotective; cytostatic; antidiabetic; vulnery; STE20: protein kinase; STIK2; STIK3; STIK4; STIK5; STIK6; STIK7; ZC1, ZC2, ZC3, ZC4, KMS2, SUDU1, SUDU3, GEX2, PAK4, PAK5; antagonist; antibody; gene therapy; rheumatoid arthritis; atherosclerosis; asthma; inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis; rhinitis; autoimmunity; organ transplantation; multiple sclerosis; myocardial infarction; cardiovascular disease; stroke; renal failure; oxidative stress-related neurodegenerative disorder; Parkinson's disease; amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy; ischemic disorder; inflammation; diabetes mellitus; fibrosis; miosis; meangial disorder; growth regulation; wound healing; T cell activation; immunosuppressant. | |
| OS | XX | Homo sapiens. | |
| XX | XX | W09953036-A2. | |
| PN | XX | 21-OCT-1999. | |
| PD | XX | 13-APR-1999; 99MO-US08150. | |
| PE | XX | 14-APR-1998; 9805-0081784. | |
| PR | XX | (SUGEN-) SUGEN INC. | |
| PA | XX | Plowman G, Martinez R, Whyte D; | |
| PI | XX | WPI; 1999-611301/52. | |
| DR | XX | N-PSDB; 240484. | |
| DR | XX | | |
| PT | XX | Novel kinase-related polypeptides used for the diagnosis and treatment of kinase-related diseases and disorders | |
| PS | XX | Claim II; Page 274-278; 387pp; English. | |
| XX | XX | This sequence represents a novel STE20-related protein kinase. The invention relates to nucleic acid molecule encoding a kinase polypeptide selected from STIK2, STIK3, STIK4, STIK5, STIK6, STIK7, ZC1, ZC2, ZC3, ZC4, KMS2, SUDU1, SUDU3, GEX2, PAK4 and PAK5. The proteins are used to identify agonists and antagonists, and to raise antibodies. The polynucleotides are useful in gene therapy protocols. The polynucleotides, polypeptides, antibodies, antagonists and agonists may be used to treat diseases such as immune-related disorders and diseases (e.g. rheumatoid arthritis, atherosclerosis, chronic inflammatory bowel disease (e.g. Crohn's disease), asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplantation, chronic inflammatory | |

CC plevia/disease, multiple sclerosis, organ transplantation, myocardial
CC infarction, cardiovascular disease, stroke, renal failure, oxidative
CC stress-related neurodegenerative disorders (e.g. amyotrophic lateral
CC sclerosis, Parkinson's disease and Leigh syndrome), cancer,
CC cardiomyopathies, ischemic disorders, inflammatory disorders, diabetes
CC mellitus, fibrotic and mesangial disorders. The proteins may also be
CC useful for cell growth regulation (e.g. in wound healing), T cell
CC activation, mitosis control, and as immunosuppressants.

| | | | |
|----|----------|------|-----|
| XX | Sequence | 1297 | AA; |
| SQ | | | |

| | | | | |
|-----------------------|-------|-----------------|-------|---------------|
| Query Match | 9.5% | Score 124.5 | DB 20 | Length 1297 |
| Best Local Similarity | 25.0% | Pred. No. 0.032 | | |
| Matches | 54 | Conservative | 43 | Mismatches 64 |
| | | | | Indels 55 |
| | | | | Gaps 10 |

```

0Y 21 LDERKKROEEMEXVKRKEDEDECPEYVODPRSLYERLOJOKRKO--OEVEDQFKRKNV 79
    | | | | | : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 laeqrkrlleeqkeqrtrleeqqrreke-----lrkqgeqrtrhyeeqmreeer 403
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 80 RGLDEDETNFLDEVSROOELEKORREBEL-KELKEYRNMLKKGISOEKKKVEKK--- 135
    | : : : : : | : | : | : | : | : | : | : | : : : : | | |
Db 404 rraehq-----eykrng--leeqrgertrqqrkgerydylvsqmqrgqrprekpply 456
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 136 -----LTVKPIETKRSQAKKLAGAVKHK-----SSSGNS 167
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 hykegmspsekpwakeveeestrlnrgs--spamphkvannrldgpnlprsesfsisgyq 514
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 168 VKRLKRPDEPPDKNOEPS--SKRLGNTSLSGPSIH 201
    | : : : : : | : | : | : | : | : | : | : | : : : : | | |
Db 515 partppmlrlpvdq-qdplnlvavksgqpalatassgvh 549
    | : : : : : | : | : | : | : | : | : | : | : : : : | | |

```

Search completed: February 16, 2001, 17:05:33
Job time: 27 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 17:05:07 ; Search time 15.5 seconds
(without alignments)
1112.693 Million cell updates/sec

Title: US-09-602-597-2

Perfect score: 1306

Sequence: 1 MDGGDDGLIKKRVSEAE.....NATGKIYSSIFRTNPLEAP 254

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

PIR66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-----------------------------|
| 1 | 302.5 | 23.2 | 219 | 2 | T19438 hypothetical prote |
| 2 | 169.5 | 13.0 | 241 | 2 | T48006 hypothetical prote |
| 3 | 158 | 12.1 | 1240 | 2 | S52734 hypothetical prote |
| 4 | 153 | 11.7 | 707 | 2 | S60588 hypothetical prote |
| 5 | 151 | 11.6 | 649 | 2 | JN0809 drebrin A - rat |
| 6 | 151 | 11.6 | 651 | 2 | T14763 drebrin E (clone g |
| 7 | 149.5 | 11.4 | 312 | 2 | T25994 hypothetical prote |
| 8 | 149 | 11.4 | 1087 | 2 | T30330 hypothetical prote |
| 9 | 146 | 11.2 | 292 | 2 | JE0233 gel-solin-related p |
| 10 | 146 | 11.2 | 451 | 2 | G70241 tropomyosin-I - scal |
| 11 | 142.5 | 10.9 | 1178 | 2 | S78475 hypothetical prote |
| 12 | 142 | 10.9 | 772 | 2 | T50463 mannosylphosphoryl |
| 13 | 142 | 10.9 | 1407 | 1 | S28589 trichomyalin - C |
| 14 | 141.5 | 10.8 | 466 | 2 | T22141 hypothetical prote |
| 15 | 141 | 10.8 | 1027 | 2 | T46481 hypothetical prote |
| 16 | 139.5 | 10.7 | 564 | 2 | B43776 drebrin EI - chick |
| 17 | 139 | 10.6 | 1233 | 2 | T30989 serine/threonine p |
| 18 | 138.5 | 10.6 | 335 | 2 | T33457 hypothetical prote |
| 19 | 138 | 10.6 | 385 | 2 | T19201 hypothetical prote |
| 20 | 137 | 10.5 | 1359 | 2 | T34036 drebrin - chicken |
| 21 | 137 | 10.5 | 1359 | 2 | T34036 hypothetical prote |
| 22 | 136.5 | 10.5 | 678 | 2 | A54514 glutamic acid-rich |
| 23 | 135 | 10.3 | 729 | 2 | T50989 hypothetical prote |
| 24 | 135 | 10.3 | 777 | 2 | T21048 hypothetical prote |
| 25 | 133 | 10.2 | 1526 | 2 | A45605 mature-parasite-in |
| 26 | 132.5 | 10.1 | 749 | 2 | I37356 epithelial microtu |
| 27 | 132.5 | 10.1 | 839 | 2 | I50590 class II INCENP pro |
| 28 | 132.5 | 10.1 | 1737 | 2 | A59235 unconventional myo |
| 29 | 132 | 10.1 | 405 | 2 | T16922 hypothetical prote |

| | | | | | |
|----|-------|------|------|---|---------------------------|
| 30 | 132 | 10.1 | 2058 | 2 | A59267 myosin X - human |
| 31 | 131.5 | 10.1 | 517 | 1 | A39038 I-caldesmon, nonmu |
| 32 | 131.5 | 10.1 | 1888 | 1 | A45973 trichomyalin - hum |
| 33 | 131 | 10.0 | 791 | 2 | T24435 hypothetical prote |
| 34 | 131 | 10.0 | 1038 | 2 | T02634 rep protein homolo |
| 35 | 130.5 | 10.0 | 1403 | 2 | T11583 probable translati |
| 36 | 130 | 10.0 | 241 | 2 | S44893 protein - |
| 37 | 130 | 10.0 | 550 | 2 | A46419 trophoblast-endoth |
| 38 | 130 | 10.0 | 695 | 2 | T54325 gene XE7 protein - |
| 39 | 129.5 | 9.9 | 665 | 2 | B71609 hypothetical prote |
| 40 | 129.5 | 9.9 | 2052 | 2 | T18519 myosin X - bovine |
| 41 | 129 | 9.9 | 812 | 2 | S43604 R0755.1 protein (c |
| 42 | 128 | 9.8 | 805 | 2 | E70474 translation initia |
| 43 | 128 | 9.8 | 1233 | 2 | I54383 chromosome segrega |
| 44 | 128 | 9.8 | 1997 | 2 | F71607 DNA helicase II BR |
| 45 | 127.5 | 9.8 | 301 | 2 | T33068 hypothetical prote |

ALIGNMENTS

```
RESULT 1
T19438
hypothetical protein C25A1.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T19438
R:Moritome, B.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19124
A:Accession: T19438
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-219 <MIL>
A:Cross-references: EMBL:Z81038; PIDN:CAB02762.1; GSPDB:GN00019; CESP:C25A1.1
A:Experimental source: clone C25A1
C:Genetics:
A:Gene: CESP:C25A1.1
A:Map position: 1
A:Introns: 33/1; 66/1; 151/2

Query Match 23 2%; Score 302.5; DB 2; Length 219;
Best Local Similarity 36.4%; Pred. No. 816-11;
Matches 82; Conservative 36; Mismatches 70; Indels 37; Gaps 8;

QY 15 FVSEALDERRRKROEWEKRPDEECPEEYVPSLYERLOKDRKQOEYEOFR 74
|||:||||:| ||| |||:|||| | | :||:|: | | | | |
Db 5 FVSTSELDEKKARQAMERIKRPDTATLVPEYCNKTLFEDLKNKKDAKQLEIDAKK 64

QY 75 FKNMVRGLDEDTNPLDEVSROQELIEKORREBELKELKRYNNLKKVGISQF--NKREV 132
|||||:||||:| | :|:|:|:| | | | | | | | | | |
Db 65 LKNMVRGIDEDSVFLSELDTSTKRVV-KMRMKREDEL-----IKELAVTQHLAANGS 117

QY 133 EKKLYTKPIETK----NKPSQAKLLAGVYKKSSEGSNSYKRLKPPPEPDKNQOESCK 188
|||||:||||:| | :|:|:|:| | | | | | | | | | |
Db 118 SSRFLIKPSTSKVLGPPKSKQAALFLSTAIKRTST-----EKKRQE----- 159

QY 189 SLGNTSLSGPSIHCPAAVCGILPGIAGVSGSDSES--SSDSE 231
: : : | | | | | | | | | | | | | | | | | | |
Db 160 DVVSSKVSKEP-----FYIKQIGALQALCEYPSDSDESADSSDSE 200

RESULT 2
T48006
hypothetical protein T17J13.100 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48006
R:Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Lemke, K.; Maye
submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24482
A:Accession: T48006
```


Db 896 EKVEENETVNEVEVG-IIVSPSEKVEA-----NSSSTISSPENEGSV 940
 QY 235 NATGK 239
 Db 941 SVKDK 945

RESULT 9

JE0233
 troponin-I - scallop (*Chlamys nipponensis*)
 C:Species: *Chlamys nipponensis* (Japanese scallop)
 C:Date: 05-Dec-1998 #sequence, revision 05-Dec-1998 #text, change 17-Mar-2000
 C:Accession: JE0233
 R:Tanaka, H.; Ojima, T.; Nishita, K.
 J. Biochem. 124, 304-310, 1998
 A:Title: Amino acid sequence of troponin-I from Akazara scallop striated adductor muscle
 A:Reference number: JE0233; MUID:98351986
 A:Accession: JE0233
 A:Molecule type: protein
 A:Residues: 1-292 <TAB>
 C:Superfamily: troponin T

Query Match 11.2%; Score 146; DB 2; Length 292;
 Best Local Similarity 28.1%; Pred. No. 0.09;
 Matches 47; Conservative 26; Mismatches 62; Indels 32; Gaps 4;

QY 18 EAELDERKKRQOEWEKRPEDPECEPEYDPRSLYERQOEKDRKQOEYEDFKKN 77
 Db 55 DAEEARRRQOEDEEAAARAEE-----YNRQOEELRRQOEERQORREE 102
 QY 78 MVRLLDETFNLDVSRQOELEKORREELKELEYRNLLKVV--GISQENKKEVEK 134
 Db 103 QRQOEERERLRER-----EEQEREERARRAEQKKKKKGGLGSLPEKKMLKK 154
 QY 135 KLYKPIETKPKFSQAKLLAGAVHKSSGNSVYKRLKPDPEPDDKN 181
 Db 155 LIMOKAED-----LANEAKAKAEKKEKYINDLVKPFSTGKD 192

RESULT 10

G70241
 hypothetical protein Bb116 - Lyme disease spirochete plasmid I/1p28-4
 C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence, revision 13-Feb-1998 #text, change 08-Oct-1999
 C:Accession: G70241
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.
 A:Reference number: A70100; MUID:98065943
 A:Accession: G70241
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-451 <KLE>
 A:Cross-references: GB:AE000789; NID:g2650079; PID:AAC66203.1; PID:g2690100; TIGR:BB116
 A:Experimental source: strain B31
 C:Genetics:
 A:Genome: plasmid

Query Match 11.2%; Score 146; DB 2; Length 451;
 Best Local Similarity 25.1%; Pred. No. 0.14; DB 2; Length 451;
 Matches 46; Conservative 46; Mismatches 77; Indels 14; Gaps 4;

QY 13 KRFSSEALDERRRRQOEWEKRPEDPECEPEYDPRSLYERQOEKDRKQOE--- 68
 Db 44 KQKSKPRTEELKKQOE-BELKKKQOEELKKQOEELKKQOEELKKQOEELKK 102
 QY 69 --YEOFKFNWVGLEDDETFLDEVSRQOELEKORREELKELEYRNLLKVGIS 125

Db 103 KKOEEELKKQOEELKKQOEELKKQOEELKKQOEELKKQOEELKKQOE 161
 QY 126 QENK-----EVEKRLVYKPIETKPKFSQAKLLAGAVHKSSGNSVYKRLKPDPEPDDK 180
 Db 162 EELKKQOEELKKQOEELKKQOEELKKQOEELKKQOEELKKQOEELKKQOEELKK 221

QY 181 NQE 183
 Db 222 QOE 224

RESULT 11

S78475
 mannosylphosphorylation protein MNNA - Yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: protein YKL200c; protein YKL201c
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 11-Mar-1998 #sequence, revision 17-Apr-1998 #text, change 05-Jan-1998
 C:Accession: S78475; S38037; S38038
 R:Odani, T.; Shima, Y.; Yoshitani, J.
 submitted to the EMBL data library, January 1996
 A:Description: Cloning and Analysis of the MNNA Gene Required for Phosphorylation of
 A:Reference number: S78475
 A:Accession: S78475
 A:Molecule type: DNA
 A:Residues: 1-1178 <ODA>
 A:Cross-references: EMBL:D83006; NID:g1752735; PID:d1012343; PID:g1752736; MIPS:YKL20
 R:Ma e Silva, A.; Bossier, P.; Vilela, C.; Fernandes, L.; Soares, H.; Guerreiro, P.
 submitted to the Protein Sequence Database, March 1994
 A:Reference number: S38024
 A:Accession: S38037
 A:Molecule type: DNA
 A:Residues: 121-249, 'A10LQ', 255, 'MALLRD', 262-390, 'LRISNS', 398-515, 'LG' <MAI>
 A:Cross-references: EMBL:Z28200; NID:g486355; PID:g486356
 A:Experimental source: strain S288C
 A:Note: this sequence has been revised in reference S78475
 A:Accession: S38038
 A:Molecule type: DNA
 A:Residues: 640-1178 <MAV>
 A:Cross-references: EMBL:Z28201; NID:g486357; PID:g486359; MIPS:YKL201c
 A:Experimental source: strain S288C
 A:Note: this sequence has been revised in reference S78475
 A:Note: this was assumed to be the complete sequence of protein YKL201c
 C:Genetics:
 A:Gene: MNNA
 A:Cross-references: MIPS:YKL201c
 A:Map position: 11L
 A:Function:
 A:Description: required for phosphorylation of N-linked oligosaccharides
 C:Keywords: transmembrane protein
 F:2844/Domain: transmembrane #status predicted <TM>

Query Match 10.9%; Score 142.5; DB 2; Length 1178;
 Best Local Similarity 27.3%; Pred. No. 0.63;
 Matches 51; Conservative 38; Mismatches 61; Indels 37; Gaps 6;

QY 6 DGNLIKKRFSSEALDERRRRQOEWEKRPEDPECEPEYDPRSLYERQOE 61
 Db 1018 DKDPIIYEDYAVAKLLERKKRREKKKEEKKKEE-----EKKRKE 1070
 QY 62 KDRKQOEYEQDFKFNWVGLEDDETFLDEVSRQOELEKORREELKELEYRNLLK 1
 Db 1071 EEEKKKEEKKKE-----EEKKKOEKEEKKKEEKKKEEKKKEGCKKN 1117
 QY 122 VGISOENKKEVEKRLVYKPIETKPKFSQAKLLAGAVHKSSGNSVYKRLKPDPEPDDKN 181
 Db 1118 E-DEENKKNDEKKKKEEKKKEE-----KNKKE--DEKKKQOEKEEKKKN 1164
 QY 182 QEPSSCK 188

Db 1165 EEEERKK 1171

RESULT 12

150463 protein kinase - chicken

C:Species: Gallus gallus (chicken)

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999

C:Accession: I50463

R:Li, H.; Grenet, J.; Valentine, M.; Lahti, J.M.; Kidd, V.J.

Gene 153, 237-242, 1995

A:Title: Structure and expression of chicken protein kinase PITSRE-encoding genes.

A:Reference number: I50463; MUID:95180725

A:Accession: I50463

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-772 <LIX>

A:Cross-References: EMBL:U16656; NID:g571458; PIDN:AAA67037.1; PID:g571460

C:Genetics:

A:Gene: cPITSRE

A:Introns: 37/3; 76/2; 118/1; 164/2; 208/1; 249/1; 283/1; 329/1; 351/1; 407/3; 438/1; 47

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

C:Keywords: ATP

F:413-668/Domain: protein kinase ATP-binding motif

F:421-429/Region: protein kinase ATP-binding motif

Query Match 10.9%; Score 142; DB 2; Length 772;

Best Local Similarity 23.9%; Pred. No. 0.43;

Matches 63; Conservative 50; Mismatches 79; Indels 72; Gaps 12;

QY 3 GGDGDLIIK-----KRFVSEALDERRRKROE-----EMEKVKRPEDPECEPEVD 50

Db 77 GEEDDSLAIKPPQOMKKEKTEHHRKRRHSHSNAVKERK-----HE 131

QY 51 PRSLRYEOKRQOEYEEQFKFNMYGLDETNFLDEVSROE--LIEKORREE 108

Db 132 RR---KRHEEDOKARREMRQ--KRREMARHSRRDRLEQERREKIRKQOEQRE 187

QY 109 LKELKRYNNKLVGISOENKKEVE-----KLTVPKPIETKNNFSQAKLLAGAV 157

Db 188 OKERERRAEERRK---EREARRVSAHHTVREYGDVKMRP-----WSRSPLRQORD 238

QY 158 KHKSSSGSVKRLKRP---DPEPD-----DKNOEPSSCKSLGTSISGPSIHPSAAYCI 209

Db 239 KPEQASRKRPVKEEKPEERDPLSDLODISDEKRTSAAKMSLQASG----- 286

QY 210 GILPGLGAYSGSSSDSSDSEGT 233

Db 287 -----SGSEEEFESESSSEGS 301

RESULT 13

S28589

trichohyalin - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999

C:Accession: S28589

R:Flitz, M.J.; Rogers, G.E.

submitted to the EMBL Data Library, December 1992

A:Description: Examination of the gene encoding rabbit trichohyalin.

A:Reference number: S28589

A:Accession: S28589

A:Molecule type: DNA

A:Residues: 1-1407 <FIE>

A:Cross-References: EMBL:Z19092; NID:g1746; PIDN:CAA79519.1; PID:g1747

C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh

covalent modifications to this protein include conversion of arginine to citrulline and

C:Genetics:

A:Introns: 46/3

C:Superfamily: trichohyalin; calmodulin repeat homology

C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat

F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 10.9%; Score 142; DB 1; Length 1407;

Best Local Similarity 26.9%; Pred. No. 0.81;

Matches 49; Conservative 39; Mismatches 48; Indels 46; Gaps 8;

QY 13 KRFSAEALDERRRK-----QEEMKY-----RKPEDECEPEYDPSRYERI 58

Db 215 KRELREEF-QQRRERQERHALQEEEQLLRORRRRREPREQQLRLELEIRERQRL 273

QY 59 OEQDKRQO-----EYEB--QFKFNMYRGIDEETNFLDEVSROELIEKORREEL 109

Db 274 EDERERQQLRREQRLEGERERQQLRLELEIRERQRLDEGERERQRLDEGERERQQL 333

QY 110 K-ELKRYNNKLVGISOENKKEVEKLVKPIETKNNFSQAKLLAGAVKHSSEGSNV 168

Db 334 KRELREIRERQER--LEQERRE-----QLLAEEVREQARREGESL 372

QY 169 KR 170

Db 373 TR 374

RESULT 14

T22141

hypothetical protein F43G9.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T22141

R:Kershaw, J.

submitted to the EMBL Data Library, September 1996

A:Reference number: Z19523

A:Accession: T22141

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-466 <MLI>

A:Cross-References: EMBL:Z79755; PIDN:CAB02103.1; GSPDB:GN00019; CESP:F43G9.10

A:Experimental source: clone F43G9

C:Genetics:

A:Gene: CESP:F43G9.10

A:Map position: 1

A:Introns: 53/2; 316/3; 407/3

Query Match 10.8%; Score 141.5; DB 2; Length 466;

Best Local Similarity 25.5%; Pred. No. 0.27;

Matches 50; Conservative 42; Mismatches 69; Indels 35; Gaps 8;

QY 12 KRFVSEALDERRRKROEEMKVKRPEDPECEPEYVPRSLYERLOQRKQOEYEE 71

Db 80 RRRRESDEDRRRHRRHEDYGRROVERPEVL-GKVEDESSSENEQSESEDEKQERRE 138

QY 72 OFKFNMY-----RGIDEETNFLDEVSROELIEKOR-----REELKELKRYRN 117

Db 139 RARMRLELHNNRREDEQ---DSAESDEDFERRRQMLDRAKRREELK--REIKE 193

QY 118 NKKVGIQENKKEVEK-----LTVKPIETKNNFSQAKLLAGAVKHSSEGSNV 167

Db 194 ELEDDEVEEESESESEDEDDDPVPLKPIETFRK--DRITLQEAKEKEKE---I 248

QY 168 VKRLKPDPEPDQKOE 183

Db 249 LKRIEDEKRAEERKRE 264

RESULT 15

T46481

hypothetical protein DKFp434A025.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T46481

R:Duesterhoft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23035
 A:Accession: T46481
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1027 <AAA>
 A:Cross-references: EMBL:AL137755
 A:Experimental source: adult testis; clone DKFzp434A025
 C:Genetics:
 A:Note: DKFzp434A025.1

Query Match 10.88; Score 141; DB 2; Length 1027;
 Best Local Similarity 23.3%; Pred. No. 0.66;
 Matches 60; Conservative 36; Mismatches 79; Indels 82; Gaps 10;

```

QY 21 LDERRRRQOE---EMEVRKKPEDEECPEEVDPRSL-----54
   :| | | | | | | | | | | | | | | | | | | | | | | |
DB 45 IDRTKKKRGKDETEYSGSEEEVPEQEGEPSSIVNPGESTLRDPLRLQENKE 104
   :| | | | | | | | | | | | | | | | | | | | | | | |
QY 55 -----YERLQEQKDRKQOEYEQF-----KFNWVRGLDEDETNFLDEYSRQOE 98
   :| | | | | | | | | | | | | | | | | | | | | | | |
DB 105 RSEALRQQLQEQQLRQOEYKRLAERQKRTEQCKEQRRRLQEQOR-EREARRQOE 163
   :| | | | | | | | | | | | | | | | | | | | | | | |
QY 99 LIEKQRRREE--LKELEKYNNTLKVGISQENKKEVEKKLTV--KPIETKKEF-----147
   :| | | | | | | | | | | | | | | | | | | | | | | |
DB 164 REQRRREOEKERRRLELEERRRRRRAEERKRVREQOYIRQLEEQRHLEVLQ 223
   :| | | | | | | | | | | | | | | | | | | | | | | |
QY 148 -----SOAKLLAGAVKHSSESGNSVKRLKP-----PEPDDKNQEPSS 186
   :| | | | | | | | | | | | | | | | | | | | | | | |
DB 224 QLQEQAMLL--ECRWREMEERHQAERLQRLQEQQAYLLSLQHDHRRRPHQSQPPPP 281
   :| | | | | | | | | | | | | | | | | | | | | | | |
QY 187 CKSLGNTSLSGPSIICP 203
   :| | | | | | | | | | | | | | | | | | | | | | | |
DB 282 -----QQERSKKPSFHAP 293
   :| | | | | | | | | | | | | | | | | | | | | | | |
  
```

Search completed: February 16, 2001, 17:06:34
 Job time: 87 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 17:05:07 ; Search time 10.27 Seconds

(Without alignments)
798.705 Million cell updates/sec

Title: US-09-602-597-2

Sequence: 1 MDGDDGNLIKKRFVSEAE.....NATKIVSSIFRTNLEAP 254

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|---------------|--------------------|
| 1 | 158 | 12.1 | 1240 | 1 YN1_YEAST | P53935 saccharomyc |
| 2 | 153 | 11.7 | 707 | 1 DREB_RAT | Q07266 rattus norv |
| 3 | 151 | 11.6 | 649 | 1 DREB_HUMAN | Q16643 homo sapien |
| 4 | 142.5 | 10.9 | 1178 | 1 MNNA_YEAST | P36044 saccharomyc |
| 5 | 142 | 10.9 | 1407 | 1 TRHY_RABIT | P37709 oryctolagus |
| 6 | 137 | 10.5 | 652 | 1 DREB_CHICK | P18302 gallus gall |
| 7 | 136.5 | 10.5 | 678 | 1 GARP_PLAIF | P13816 plasmodium |
| 8 | 132 | 10.1 | 405 | 1 TRT_CAEEL | Q27371 caenorhabd |
| 9 | 131.5 | 10.1 | 1898 | 1 TRHY_HUMAN | Q07283 homo sapien |
| 10 | 130.5 | 10.0 | 1403 | 1 YDF3_SCHPO | Q10475 schizosacch |
| 11 | 130 | 10.0 | 241 | 1 YOB7_CAEEL | P34623 caenorhabd |
| 12 | 130 | 10.0 | 550 | 1 BLSA_HUMAN | Q02832 homo sapien |
| 13 | 130 | 10.0 | 695 | 1 XE7_HUMAN | Q02040 homo sapien |
| 14 | 128 | 9.8 | 805 | 1 IF2_AQUAE | Q07825 aquifex aeo |
| 15 | 127.5 | 9.8 | 600 | 1 LAM2_CHICK | P14732 gallus gall |
| 16 | 127.5 | 9.8 | 793 | 1 CALD_HUMAN | Q05682 homo sapien |
| 17 | 126.5 | 9.7 | 407 | 1 M21_STRPY | P50468 streptococc |
| 18 | 126.5 | 9.7 | 877 | 1 INCE_CHICK | P53352 gallus gall |
| 19 | 126.5 | 9.7 | 2116 | 1 MY52_DICDI | P08799 dictyostell |
| 20 | 126 | 9.6 | 410 | 1 YG31_YEAST | P53866 saccharomyc |
| 21 | 125.5 | 9.6 | 1959 | 1 MY5N_CHICK | P14105 gallus gall |
| 22 | 125 | 9.6 | 897 | 1 EP15_MOUSE | P42567 mus musculu |
| 23 | 125 | 9.6 | 1549 | 1 TRHY_SHEEP | P22793 ovis aries |
| 24 | 124.5 | 9.5 | 657 | 1 KNOB_PLAIFN | P06719 plasmodium |
| 25 | 123.5 | 9.5 | 816 | 1 HUNB_DROVI | P13361 drosophila |
| 26 | 123 | 9.4 | 1234 | 1 PIP3_MOUSE | P51432 mus musculu |
| 27 | 122.5 | 9.4 | 771 | 1 CALD_CHICK | P12957 gallus gall |
| 28 | 122.5 | 9.4 | 1085 | 1 YARA_SCHPO | Q09663 schizosacch |
| 29 | 122.5 | 9.4 | 1220 | 1 IF2P_HUMAN | O60841 homo sapien |
| 30 | 122.5 | 9.4 | 1391 | 1 MY52_DROHY | Q08596 drosophila |
| 31 | 122 | 9.3 | 412 | 1 YNP1_CAEEL | P34554 caenorhabd |
| 32 | 121.5 | 9.3 | 304 | 1 CEC1_CAEEL | P34518 caenorhabd |
| 33 | 121.5 | 9.3 | 634 | 1 KNOB_PLAIFG | P09346 plasmodium |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 34 | 121.5 | 9.3 | 887 | 1 YLX8_CAEEL | P46504 caenorhabd |
| 35 | 121.5 | 9.3 | 1023 | 1 GLT1_DROME | P33438 drosophila |
| 36 | 121.5 | 9.3 | 1118 | 1 UBP8_HUMAN | P40818 homo sapien |
| 37 | 121.5 | 9.3 | 1940 | 1 MYSE_RAT | P12847 rattus norv |
| 38 | 120.5 | 9.2 | 848 | 1 NFW_MOUSE | P08553 mus musculu |
| 39 | 120.5 | 9.2 | 1940 | 1 MYSE_HUMAN | P11055 homo sapien |
| 40 | 120 | 9.2 | 407 | 1 IE68_HSVSA | Q01042 herpesvirus |
| 41 | 119.5 | 9.2 | 396 | 1 TRT_DROME | P19351 drosophila |
| 42 | 119.5 | 9.2 | 1790 | 1 USO1_YEAST | P25386 saccharomyc |
| 43 | 119.5 | 9.2 | 1960 | 1 MTSN_HUMAN | P35579 homo sapien |
| 44 | 119 | 9.1 | 285 | 1 TPM1_DROME | P06754 drosophila |
| 45 | 118.5 | 9.1 | 542 | 1 TUL1_HUMAN | Q00294 homo sapien |

ALIGNMENTS

| RESULT | 1 | STANDARD | PRT | 1240 AA. |
|-----------------------|--|------------------|-------|-----------------------------------|
| YN1_YEAST | YN1_YEAST | | | |
| AC | P53935 | | | |
| DT | 01-OCT-1996 (Rel. 34, Created) | | | |
| DT | 01-OCT-1996 (Rel. 34, Last sequence update) | | | |
| DT | 01-OCT-1996 (Rel. 34, Last annotation update) | | | |
| DE | HYPOTHETICAL 141.5 KDA PROTEIN IN YPT53-RHO2 INTERGENIC REGION. | | | |
| GN | YN1091W OR N2231. | | | |
| OS | Saccharomyces cerevisiae (Baker's yeast). | | | |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales; | | | |
| OC | Saccharomycetaceae; Saccharomycetes. | | | |
| RN | [1] | | | |
| RC | SEQUENCE FROM N.A. | | | |
| RP | SPRAIN-5288C / FY1679; | | | |
| RX | MEDLINE-96367601; PubMed-8771715; | | | |
| RA | Garcia-Cantalejo J.M., Boskovic J., Jimenez A.; | | | |
| RT | "Sequence analysis of a 14.2 kb fragment of Saccharomyces cerevisiae | | | |
| RT | chromosome XIV that includes the ypt53, tRNAleu and gsr m2 genes and | | | |
| RT | four new open reading frames." | | | |
| RL | Yeast 12:599-608(1996). | | | |
| CC | - SIMILARITY: TO S.POMBE SPAC29B6.10C. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL: X85811; CA59826.1; - | | | |
| DR | EMBL: Z71367; CA95967.1; - | | | |
| DR | HSSP: P03069; IYSA. | | | |
| DR | SDD: S0005035; YNL091W. | | | |
| KW | Hypothetical protein. | | | |
| FT | DOMAIN 756..761 | | | |
| SO | SEQUENCE 1240 AA; 141513 MW; 3FE9D26582D5778 CRC64: 4 | | | |
| Query Match | 12.1%; | Score 158; | DB 1; | Length 1240; |
| Best Local Similarity | 23.8%; | Pred. No. 0.051; | | |
| Matches | 54; | Conservative | 48; | Mismatches 79; Indels 46; Gaps 8; |
| QY | 6 DGNLIKKRFVSEAELEDERKKRQOEWEKYPPEDEPECEVEYDPSLYERLOEQKDK 65 | | | |
| DB | 621 DNN---RLKLQLE-EEKKRRKEKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 669 | | | |
| QY | 66 QOEYEEQF-----KFNKNVRGDEDEETNLEDEVSQOEIEKOR----- 104 | | | |
| DB | 670 REEKKELKLEERERKKRQAKKYEAKKKKKDEKKRRLLEQQAKREKQEKORQKE 729 | | | |
| QY | 105 -----REEELKEKEKRRNNLKVGISOENKEVEKKLTVPRIETKNNFSQAKLAVKH 159 | | | |
| DB | 730 ELKRRKEKKRIRE-QRLQEKLOKKEKEEERQRLIADALRKQKLNBEQTSANILSA 788 | | | |

QY 160 KSSSGNSVKKRLKDPEDDKN-OEPSSC-----KSLGNTSLSGP 198
 Db 789 KPTENGNGNPVSSOSHNMNTNYOEDNCSINDEILKMNVSVAASKP 835

RESULT 2

REB_RAT STANDARD: PRT: 707 AA.
 ID DREB_RAT
 AC 007266;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DREBRIN A.
 GN DBN1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WISTAR; TISSUE=BRAIN, AND HIPPOCAMPUS;
 RX MEDLINE=92305233; PubMed=1611026;
 RA Shirao T., Obata N., Obata K.;
 RT "Cloning of drebrin A and induction of neurite-like processes in
 drebrin-transfected cells."
 RL Neuroreport 3:109-112(1992).
 CC -1- FUNCTION: DREBRINS MIGHT PLAY SOME ROLE IN CELL MIGRATION,
 CC EXTENSION OF NEURONAL PROCESSES AND PLASTICITY OF DENDRITES,
 CC RESPECTIVELY. BINDS ACTIN.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- ALTERNATIVE PRODUCTS: MULTIPLE FORMS OF DREBRINS RESULT FROM
 CC ALTERNATIVE SPLICING OF THE SINGLE DREBRIN GENE DURING NEURAL
 CC DEVELOPMENT.
 CC -1- TISSUE SPECIFICITY: BRAIN NEURONS.
 CC -1- MISCELLANEOUS: DREBRINS ARE CLASSIFIED INTO TWO FORMS OF THE
 CC EMBRYONIC TYPE (E1 AND E2) AND ONE FORM OF THE ADULT TYPE (A). THE
 CC TIME COURSE OF THEIR APPEARANCE ARE DIFFERENT FROM EACH OTHER.
 CC THEIR STRUCTURES ARE CLOSELY RELATED.
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X59267; CAA1957.1; -
 DR INTERPRO: IPR002108; -
 DR PFM: PF00241; Cofilin_ADF.1.
 KW Actin-binding; Brain; Neurone; Alternative splicing.
 SQ SEQUENCE 707 AA; 77471 MW; B5279BF6EB7B80AA CRC64;

Query Match 11.7%; Score 153; DB 1; Length 707;
 Best Local Similarity 25.7%; Pred. No. 0.053;
 Matches 61; Conservative 47; Mismatches 81; Indels 48; Gaps 10;

QY 17 SEAELEDRKRRKROEWEVVRKPEDEPCPEVYVPRSLYERLOQOKRQOEYEQFK 76
 Db 166 TDAVEMKRRINRQEWQAKR---EEELRKEDEKKALDARLRFEQEMERQDQ----- 218
 QY 77 NMVGLDEDTNPLDEVSRQOELIEKORREBELKELKEYRNLLKVGISQENKKEVEK 136
 Db 219 -----EERERRRRE---REQQ-IEEHRKQOQSLAEAEAKRRLKQOSIFGDHDEE--- 265
 QY 137 TVKPLETKNKSQAKLLAGAVKHKSSSEGSNVKRLKP-PEPDDKNQPSCKSLGNTSL 195
 Db 266 -----ESQMKKSESEV-----EAAAIIAQRPDNPREFFRQOQERVASASGSCDA 310
 QY 196 SGPSTHCPASAIVCIGILPGLGAY--SGSSDSSESSDSSES-----TINATGIYSSI 244
 Db 311 PSPFNHRGPRYC---PFIKASDSGSPSSSSSSSSSPPTPTTCHRTNLSL 363

RESULT 3

REB_HUMAN STANDARD: PRT: 649 AA.
 ID DREB_HUMAN
 AC 016643;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE DREBRIN E.
 GN DBN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=OSTEOBLAST;
 RA Fisher L.W., McBride O.W., Filipula D., Ibaraki K., Young M.F.;
 RT "Human drebrin: cDNA sequence, mRNA tissue distribution and
 RT chromosomal localization."
 RL Neurosci. Res. Commun. 14:35-42(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FETAL BRAIN;
 RX MEDLINE=94030036; PubMed=8216329;
 RA Toda M., Shirao T., Minoshima S., Shimizu N., Toya S., Uyemura K.;
 RT "Molecular cloning of cDNA encoding human drebrin E and chromosomal
 RT mapping of its gene."
 RL Biochem. Biophys. Res. Commun. 196:468-472(1993).
 CC -1- FUNCTION: DREBRINS MIGHT PLAY SOME ROLE IN CELL MIGRATION,
 CC EXTENSION OF NEURONAL PROCESSES AND PLASTICITY OF DENDRITES,
 CC RESPECTIVELY. BINDS ACTIN.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: BRAIN NEURONS. ALSO FOUND IN THE HEART,
 CC PLACENTA, SKELETAL MUSCLE, KIDNEY AND PANCREAS.
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U00802; AAA16256.1; -
 DR EMBL: D17530; BAA04480.1; -
 DR MIM: 126660; -
 DR INTERPRO: IPR002108; -
 DR PFM: PF00241; Cofilin_ADF.1.
 KW Actin-binding; Brain; Neurone.
 SQ SEQUENCE 649 AA; 71425 MW; A1730E7C5FC32D88 CRC64;

Query Match 11.6%; Score 151; DB 1; Length 649;
 Best Local Similarity 25.8%; Pred. No. 0.063;
 Matches 49; Conservative 44; Mismatches 61; Indels 36; Gaps 7;

QY 17 SEAELEDRKRRKROEWEVVRKPEDEPCPEVYVPRSLYERLOQOKRQOEYEQFK 76
 Db 166 TDAVEMKRRINRQEWQAKR---EEELRKEDEKKALDARLRFEQEMERQDQ----- 218
 QY 77 NMVGLDEDTNPLDEVSRQOELIEKORREBELKELKEYRNLLKVGISQENKKEVEK 136
 Db 219 -----EERERRRRE---REQQ-IEEHRKQOQSLAEAEAKRRLKQOSIFGDHDEE--- 265
 QY 137 TVKPLETKNKSQAKLLAGAVKHKSSSEGSNVKRLKP-PEPDDKNQPSCKSLGNTSL 195
 Db 266 -----ETHMKKSESEV-----EAAAIIAQRPDNPREFFRQOQERVASASGSCDV 310
 QY 196 SGPSTHCPSA 205
 Db 311 PSPFNHRGGS 320

```

RESULT 4
MNN4_YEAST STANDARD: PRT; 1178 AA.
ID MNN4_YEAST
AC P36044; P36043; P89095;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MNN4 PROTEIN.
GN MNN4 OR YKL200C/YKL201C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
NC Saccharomycetaceae; Saccharomyces.
RC [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97175967; PubMed=9023541;
RA Odani T., Shimizu Y., Tanaka A., Jigami Y.;
RT "Cloning and analysis of the MNN4 gene required for phosphorylation
RL of N-linked oligosaccharides in Saccharomyces cerevisiae.";
GL Glycobiology 6:805-810(1996).
[2]
RP SEQUENCE FROM N.A.
RA Maia E Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
RA Guerreiro P., Rodrigues-Pousada C.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY FUNCTION AS A POSITIVE REGULATOR FOR
CC MANNOSYLPHOSPHATE TRANSFERASE. IS REQUIRED TO MEDIATE
CC MANNOSYLPHOSPHATE TRANSFER IN BOTH THE CORE AND OUTER CHAIN
CC PORTIONS OF N-LINKED OLIGOSACCHARIDES.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: TO YEAST YJR061W.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS, ONE OF WHICH PRODUCES TWO SEPARATE ORFS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D83006; BAAL1676.1; -
DR EMBL; Z28201; CAA82046.1; -
DR EMBL; Z28200; CAA82044.1; -
DR PIR; S38037; S38037.
DR YEPD; 6530; -
DR SGD; S0001683; MNN4.
KW Transmembrane; Signal-anchor; Repeat.
FT DOMAIN 1 27 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 28 48 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT DOMAIN 49 1178 LUMENAL (POTENTIAL).
FT DOMAIN 1032 1174 ARG/GLU/LYS-RICH (HIGHLY CHARGED).
FT DOMAIN 1042 1174 17 x 8 AA TANDEM REPEATS OF K-K-K-E-E-
FT FT E-E.
FT REPEAT 1042 1049 1.
FT REPEAT 1050 1057 2.
FT REPEAT 1058 1065 3.
FT REPEAT 1066 1073 4.
FT REPEAT 1074 1081 5.
FT REPEAT 1082 1089 6.
FT REPEAT 1090 1097 7 (APPROXIMATE).
FT REPEAT 1098 1105 8.
FT REPEAT 1106 1113 9 (APPROXIMATE).
FT REPEAT 1114 1121 10 (APPROXIMATE).
FT REPEAT 1122 1129 11 (APPROXIMATE).
FT REPEAT 1130 1137 12.
FT REPEAT 1138 1144 13 (APPROXIMATE).
FT REPEAT 1145 1152 14 (APPROXIMATE).
FT REPEAT 1153 1160 15 (APPROXIMATE).
FT REPEAT 1161 1168 16 (APPROXIMATE).

```

```

FT REPEAT 1169 1174 17 (INCOMPLETE).
FT DOMAIN 37 40 POLY-LYS.
FT DOMAIN 1042 1045 POLY-LYS.
FT DOMAIN 1046 1049 POLY-LYS.
FT DOMAIN 1050 1053 POLY-LYS.
FT DOMAIN 1054 1057 POLY-LYS.
FT DOMAIN 1058 1061 POLY-LYS.
FT DOMAIN 1062 1065 POLY-LYS.
FT DOMAIN 1066 1069 POLY-LYS.
FT DOMAIN 1070 1073 POLY-LYS.
FT DOMAIN 1074 1077 POLY-LYS.
FT DOMAIN 1078 1081 POLY-LYS.
FT DOMAIN 1082 1085 POLY-LYS.
FT DOMAIN 1086 1089 POLY-LYS.
FT DOMAIN 1094 1097 POLY-LYS.
FT DOMAIN 1098 1101 POLY-LYS.
FT DOMAIN 1102 1105 POLY-LYS.
FT DOMAIN 1134 1137 POLY-LYS.
FT DOMAIN 1157 1160 POLY-LYS.
FT DOMAIN 1165 1168 POLY-LYS.
SQ SEQUENCE 1178 AA; 139380 MW; BC05DAE0AEFCB282 CRC64;

Query Match 10.9%; Score 142.5; DB 1; Length 1178;
Best Local Similarity 27.3%; Pred. No. 0.35;
Matches 51; Conservative 38; Mismatches 61; Indels 37; Gaps 6;

QY 6 DGNLIIRKRVSEALDERRRKRO---EEMVKRPDEDECEVEYDPSIYERLQEQ 61
DB 1018 DKPIIYEDYVAALKLEERKRREKKKEEKKKEE-----EKKKKE 1070
QY 62 KDRKROEYEROFKNNVRLDEDTFVLDEVSNOOLIKOREEELKEKERNMLK 121
DB 1071 EEEKKKEEKKKEE-----EKKKKEEKKKEEKKKEEKKKEEKKKEEKKK 1117
QY 122 VGISOENKKEVEKKLVKPIETKNNKFSQAKILAGAVKKSSEGSNSYKRLKPPEDDK 181
DB 1118 E--DEEKKKDEDEKKKEEKKKEE-----KNNKNE--DEKKKKEEKKKN 1164
QY 182 QEPSSCK 188
DB 1165 EEEKKK 1171

RESULT 5
TRHY_RABIT STANDARD: PRT; 1407 AA.
ID TRHY_RABIT
AC P37709;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRICHOHYALIN.
GN TRH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RA Fietz M.J., Rogers G.E.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIE NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).

```

CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIP DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -1- PIR: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z19092; CAA79519.1; -
CC PIR; S28589; S28589.
CC HSSP; P02633; IBOC.
CC INTERPRO; IPR001751; -
CC INTERPRO; IPR002048; -
CC DR PFAM; PF01023; S_100; 1.
CC DR PFAM; PF00036; eHand; 1.
CC DR PROSITE; PS00018; EF HAND; 1.
CC DR PROSITE; PS00303; S100 CABP; 1.
CC KW Repeat: Calcium-binding.
CC FT DOMAIN 1 91 SITE I (LOW AFFINITY) (POTENTIAL).
CC CA BIND 22 33 SITE II (HIGH AFFINITY) (POTENTIAL).
CC FT CA BIND 62 73 SITE II (HIGH AFFINITY) (POTENTIAL).
CC SQ SEQUENCE 1407 AA; 183781 MW; A617D2A159F12B7F CRC64;

Query Match 10.9%; Score 142; DB 1; Length 1407;
Best Local Similarity 26.9%; Pred. No. 0.46;
Matches 49; Conservative 39; Mismatches 48; Indels 46; Gaps 8;

QY 13 KRFVSEALDEKRRR-----QEWKRV-----RKPEDEPCPEVYDPRSLYERL 58
DB 215 KRELREEF-QQRRERREHERALQEEBQDLRQRWRREPRQOOLKRELEIRERQRL 273
QY 59 QEKDKRQO-----EYEE-QRFKKNVRLGDEDTNFLDEVSRQOELIEKQREBEL 109
DB 274 EQEERREQOQLRREQLRQEEERREQQLRRELEIRERQRLQEEERREQRLQEEERQOL 333
QY 110 K-ELKRYNNLNKVGISQENKKEVKLTVPRIETKNNFSQAKLLAGAVKHKSSGNSV 168
DB 334 KRELREIRERDQR-LQEEERRE-----QLLAEEVREQARERGESL 372
QY 169 KR 170
DB 373 TR 374

RESULT 6
DREB_CHICK STANDARD: PRT: 652 AA.
AC P18302; Q91358; Q91359;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DREBRINS A, E1 AND E2.
GN DBNL
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
CC OC Gallus.
CC [1]
CC RP SEQUENCE FROM N.A. (FORMS E1 AND E2).
CC RC TISSUE-BRAIN.
CC RX MEDLINE-89089137; PubMed-3208110;
CC RA Kojima N., Kato Y., Shirao T., Ohta K.;
CC "Nucleotide sequences of two embryonic drebrins, developmentally
CC regulated brain proteins, and developmental change in their mRNAs";
CC RL Brain Res. 464:207-215(1988).
CC [2]
CC RP SEQUENCE FROM N.A. (FORM A).
CC RX MEDLINE-93368392; PubMed-8361332;
CC RA Kojima N., Shirao T., Ohta K.;
CC "Molecular cloning of a developmentally regulated brain protein,
CC chicken drebrin A and its expression by alternative splicing of the
CC drebrin gene";
CC RL Brain Res. Mol. Brain Res. 19:101-114(1993).
CC -1- FUNCTION: DREBRINS MIGHT PLAY SOME ROLE IN CELL MIGRATION,
CC EXTENSION OF NEURONAL PROCESSES AND PLASTICITY OF DENDRITES,
CC RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- ALTERNATIVE PRODUCTS: MULTIPLE FORMS OF DREBRINS RESULT FROM
CC ALTERNATIVE SPLICING OF THE SINGLE DREBRIN GENE DURING NEURAL
CC DEVELOPMENT.
CC -1- TISSUE SPECIFICITY: BRAIN NEURONS.
CC -1- MISCELLANEOUS: DREBRINS ARE CLASSIFIED INTO TWO FORMS OF THE
CC EMBRYONIC TYPE (E1 AND E2) AND ONE FORM OF THE ADULT TYPE (A). THE
CC TIME COURSE OF THEIR APPEARANCE ARE DIFFERENT FROM EACH OTHER.
CC THEIR STRUCTURES ARE CLOSELY RELATED.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M36961; AAA48750.1; -
CC DR EMBL; S65296; AAB28012.1; -
CC DR EMBL; S65279; AAB28012.1; JOINED.
CC DR EMBL; S65280; AAB28012.1; JOINED.
CC DR EMBL; S65281; AAB28012.1; JOINED.
CC DR EMBL; S65288; AAB28012.1; JOINED.
CC DR EMBL; S65289; AAB28012.1; JOINED.
CC DR EMBL; S65290; AAB28012.1; JOINED.
CC DR EMBL; S65291; AAB28012.1; JOINED.
CC DR EMBL; S65292; AAB28012.1; JOINED.
CC DR EMBL; S65294; AAB28012.1; JOINED.
CC DR EMBL; S65267; CAB32515.1; -
CC DR EMBL; S65230; AAB28010.1; -
CC PIR; A43776; A43776.
CC DR INTERPRO; IPR002108; -
CC PFAM; PF00241; Cofilin_ADF; 1.
CC KW Actin-binding; Brain; Neurone; Alternative splicing.
CC FT DOMAIN 47 50 POLY-GLY.
CC FT DOMAIN 332 339 POLY-SER.
CC FT DOMAIN 542 552 POLY-GLU.
CC FT DOMAIN 643 646 POLY-GLU.
CC FT VARSPIC 317 362 MISSING (IN ISOFORM E1).
CC FT VARSPIC 307 405 MISSING (IN ISOFORM E2).
CC FT CONFLICT 552 552 E -> EE (IN REF. 1).
CC SQ SEQUENCE 652 AA; 71535 MW; 8BF72CA549833E65 CRC64;

Query Match 10.5%; Score 137; DB 1; Length 652;
Best Local Similarity 24.8%; Pred. No. 0.38;
Matches 54; Conservative 43; Mismatches 73; Indels 48; Gaps 8;

QY 17 SEALDERKRRKQEEWKEVKKPEDEPCPEVYDPRSLYERLQEKDKRQOELIEKQREBEL 76
DB 166 TDAVEMKRLNDEQWQAKK---EELRKDEERKALDARLRFQEBDEERLEQ----- 218

AC 007283;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE TRICHOHYALIN.
 GN THH OR TRHY OR THL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93280194; PubMed-7685034;
 RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
 RA Steinert P.M.;
 RT "The structure of human trichohyalin. Potential multiple roles as a
 functional EF-hand-like calcium-binding protein, a cornified cell
 envelope precursor, and an intermediate filament-associated (cross-
 linking) protein.";
 RT J. Biol. Chem. 268:12164-12176(1993).
 RL (2)
 RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE-93315897; PubMed-7686953;
 RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
 RT "Trichohyalin: a structural protein of hair, tongue, nail, and
 epidermis.";
 RT J. Invest. Dermatol. 101:655-715(1993).
 CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
 IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
 INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
 LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
 ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
 WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
 ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
 ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
 DIFFERENTIATION.
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
 THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
 THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
 THE EPIDERMIS.
 CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
 CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
 ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
 OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
 ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
 THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
 CC -1- DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
 THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
 DIFFERENT SPECIES.
 CC -1- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
 PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
 FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L09190; AAA65582.1; -
 CC PIR: A45973; A45973.
 CC DR HSSP: P02633; IBOC.
 CC DR MIM: 190370; -
 CC DR INTERPRO: IPR001751; -
 CC DR INTERPRO: IPR002048; -
 CC DR PFAM: PF01023; S_100; 1.
 CC DR PFAM: PF00036; efhand; 1.
 CC DR PROSITE: PS00018; EF_HAND; 1.

DR PROSITE; PS00303; S100_CABP; 1.
 KW Repeat; Calcium-binding.
 FT DOMAIN 1 91
 FT CA_BIND 22 33
 FT CA_BIND 62 73
 FT DOMAIN 314 390
 FT REPEAT 314 326
 FT REPEAT 327 339
 FT REPEAT 340 351
 FT REPEAT 352 364
 FT REPEAT 365 377
 FT REPEAT 378 390
 FT DOMAIN 391 444
 FT REPEAT 391 396
 FT REPEAT 397 402
 FT REPEAT 403 408
 FT REPEAT 409 414
 FT REPEAT 415 420
 FT REPEAT 421 426
 FT REPEAT 427 432
 FT REPEAT 433 438
 FT REPEAT 439 444
 FT DOMAIN 444 702
 FT REPEAT 923 1162
 FT REPEAT 923 952
 FT REPEAT 953 982
 FT REPEAT 983 1012
 FT REPEAT 1013 1042
 FT REPEAT 1043 1072
 FT REPEAT 1073 1102
 FT REPEAT 1103 1132
 FT REPEAT 1133 1162
 FT DOMAIN 1250 1849
 FT REPEAT 1752 1752
 FT CONFLICT 1794 1801
 FT CONFLICT 1857 1857
 FT CONFLICT 1880 1880
 SQ SEQUENCE 1898 AA; 247219 MW; A7AB5947FB62E31D CRC64;
 Query Match 10.1%; Score 131.5; DB 1; Length 1898;
 Best Local Similarity 25.8%; Pred. No. 2.4;
 Matches 41; Conservative 35; Mismatches 46; Indels 37; Gaps 5;
 Oy 12 KKRFSSEAL-----DERRKRQEMKVRKPED-----PECEEYDPSLYER 57
 Db 1004 EKYREBELQOEEBQLREERERKRROEWERYKKDELQOEEBQLREERERKRLOER 1063
 Oy 58 -----LQEKD-----RKQOYEEOEFKMMVRLGDEDTNFIDEVSRQOE 99
 Db 1064 EROYREBELQOEEBQLLGEERERRRQRELQRYKKBELQOEEBQLREERERKRLOER 1123
 Oy 100 ICKORREELKE-----LKEYRNLIKVGISQENKKEVE 133
 Db 1124 EROCREEELQOEEBQLREERERKRROELEROYREEE 1162
 RESULT 10
 YDF3_SCHPO STANDARD; PRT: 1403 AA.
 AC Q10475;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROBABLE EUKARYOTIC INITIATION FACTOR C17C9.03.
 GN SPAC17C9.03.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
 CC Schizosaccharomycetaceae; Schizosaccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;

RA Murphy L., McDougall R., Jones L., Simpson I., McNeill A., Harris D.,
RA Barrett B.G., Rajadream M.A., Walsh S.V.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST EUKARYOTIC INITIATION FACTOR 4F SUBUNITS P130
CC AND P150.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: 273099; CAA97349.1; -
KW Hypothetical protein; Initiation factor; Protein biosynthesis;
KW RNA-binding.
SQ SEQUENCE 1403 AA; 154034 MW; 0317EE65BE2A1E63 CRC64;

Query Match 10.0%; Score 130.5; DB 1; Length 1403;
Best Local Similarity 25.0%; Pred. No. 2;
Matches 56; Conservative 42; Mismatches 105; Indels 21; Gaps 7;

QY 17 SEAEIDRRRRROEWEKVEKPEDEPECEVYDPRSLYERLOEKDRKQOEYEQPFK 76
DB 593 AKREAENNAKREAEKAKREAEKAKREAEKAKREAEKAKREAEKAKREAEKAKRE 652
QY 77 NMVAGLDEDETNFLDEVSR--QOELIEKORREBELKELKRYNNLKVGISQENKEVER 134
DB 653 AEEAKKREAEENAKREAEKAKREAEENAKREAEKAKREAEENAKKKA--EEGSKREADK 711
QY 135 KLVYKPIETFNKKSQAKLLAGAVKHKHSSSEGSNVKRLKDPEDDDKN--QEPSSCKSLGN 192
DB 712 NPEIK-----SSAPLASSEANVDTSKOTNAT-----EPVADTKYKVKELKASEGKST 758
QY 193 TSLSGPSIHCPASAACVIGIPLGAYSGSSDSSSSDSEGTIA 236
DB 759 SLSLSPS-HETSSK--RDLLSGLESLSLKTNPKEOCELESLS 799

RESULT 11
Y087 CAEEL
ID Y087 CAEEL STANDARD; PRT; 241 AA.
AC P34623;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOTHETICAL 28.5 KDA PROTEIN ZK1236.7 IN CHROMOSOME III.
GN ZK1236.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell J., Fraser A.,
RA Fulton L., Garner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnson L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Patrillo P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roop A., Saunders D., Showmken R.,
RA Sims M., Shalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Therry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohleman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.*;
RL Nature 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: I13200; AAA28188.1; -
DR PIR: S44893; S44893.
DR MORPEP: ZK1236.7; CE00384.
KW Hypothetical protein.
SQ SEQUENCE 241 AA; 28541 MW; 610E0A6537159FA7 CRC64;

Query Match 10.0%; Score 130; DB 1; Length 241;
Best Local Similarity 25.7%; Pred. No. 0.33;
Matches 56; Conservative 39; Mismatches 63; Indels 60; Gaps 11;

QY 1 MDGQ-----DDGNLIIT-----KRFVSEALDER--RRROEWEKVKRPED 40
DB 32 LDGGAQGFYEDEGKTKIGKRAKALQAKKEKROMREYEVNERERKREERKRR---D 88
QY 41 PEBCPEEYDPRSLYERLOEKDRKQOEYEQPFKNMVGLEDDETNFLDEVSRQOELI 100
DB 89 EERAKEEA--DEKAEERLKREREREKERKEHEEYLAKKASFAIEEGT-----DAI 137
QY 101 EKORREBELKELKRYNNLKVGISQEN-----KKE-----VEKLTVPETFKNK 146
DB 138 EGGEAENLRDFVDYKTNVNVNIDELSHFGKSEDAVNRLOHFTFEEGLVGQVMDRCK 197
QY 147 F-----SOAKL--AGAVK-HKSSSEGSNVKRLK 172
DB 198 FTISDEEFAVAKFTINQGRVSIHEIADQSNLRLE 235

RESULT 12
BLSA HUMAN
ID BLSA HUMAN STANDARD; PRT; 550 AA.
AC 002832;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE B-LYMPHOCYTE ANTIGEN PRECURSOR (B-LYMPHOCYTE SURFACE ANTIGEN) (721P).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA;
RX MEDLINE=93066251; PubMed=1438229;
RA Voland J.R., Wyzzykowski R.J., Huang M., Dutton R.W.;
RT Cloning and sequencing of a trophoblast-endothelial-activated
RT lymphocyte surface protein: cDNA sequence and genomic structure.*;
RL Proc. Natl. Acad. Sci. U.S.A. 89:10425-10429(1992).
CC -1- FUNCTION: THIS PROTEIN MAY BE INVOLVED IN B-CELL ACTIVATION.
CC MAY ALSO BE INVOLVED IN SIGNAL TRANSDUCTION AND GENE REGULATION.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (PROBABLE).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ACTIVATED LYMPHOCYTES. MOST
CC VASCULAR ENDOTHELIAL AND SYNCYTIOOTROPHOBLAST.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M99578; AAA36187.1; -
DR PIR: A46419; A46419.
KW Glycoprotein; Signal; Membrane.
FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 550 B-LYMPHOCYTE ANTIGEN.
 FT DOMAIN 344 365 LEUCINE-ZIPPER.
 FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 550 AA; 64120 MW; 0298993855FB80E7 CMC64;

Query Match 10.0%; Score 130; DB 1; Length 550;
 Best Local Similarity 26.5%; Pred. No. 0.79;
 Matches 50; Conservative 41; Mismatches 56; Indels 42; Gaps 8;

QY 4 GDDGLI---IKRFRVSAEL-DEKRRKROEMEKVKKPEDECEPEVYDPRSLYERLQ 59
 DB 243 GEDGAACVACNIKVSFDSKHLSDASIKRROLERKLOE-----LEQOR 285
 QY 60 EOKDKOQOEYEQEFKMMVNGIDEDENFLDEVSRQOELIEKORREBELKELKRYNNL 119
 DB 286 EOKRREKEAEER-----QAEERKOKLELETEREKREKREKRLKROKORDRELRRNQ 338
 QY 120 KKV-GISOENKKEVEKTLVPIETKPKFSQ-----AKLT--AGAVKHSSESGNS 167
 DB 339 KKLKLEQAEQKQLOEKIKLE--ERKLLAQNRQISILIAELLISRAKAVKILREGEQKEE 396
 QY 168 VKRLKPDE 176
 DB 397 KLRLOQOE 405

RESULT 13

XE7_HUMAN STANDARD; PRT; 695 AA.

AC Q02040;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE PROTEIN XE7.
 GN XE7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RX MEDLINE-93258310; PubMed-1302606;
 RA Ellison J.W., Ramos C., Yen P.H., Shapiro L.J.;
 RT "Structure and expression of the human pseudautosomal gene XE7."
 RL Hum. Mol. Genet. 1:691-696(1992).
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
 CC SPLICING.
 CC -1- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED.
 CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L03426; AAA61304.1; -;
 DR EMBL; L03426; AAA61303.1; -;
 DR MIM; 312095; -;
 DR MIM; 465000; -;
 KM Alternative splicing.
 FT VASAPLIC 385 385 A -> L (IN SHORT ISOFORM).
 FT VASAPLIC 386 695 MISSING (IN SHORT ISOFORM).
 SO SEQUENCE 695 AA; 80709 MW; 9FFB9DF617BDC8 CRC64;

Query Match 10.0%; Score 130; DB 1; Length 695;
 Best Local Similarity 26.5%; Pred. No. 1;
 Matches 50; Conservative 41; Mismatches 56; Indels 42; Gaps 8;
 QY 4 GDDGLI---IKRFRVSAEL-DEKRRKROEMEKVKKPEDECEPEVYDPRSLYERLQ 59

DB 243 GEDGAACVACNIKVSFDSKHLSDASIKRROLERKLOE-----LEQOR 285
 QY 60 EOKDKOQOEYEQEFKMMVNGIDEDENFLDEVSRQOELIEKORREBELKELKRYNNL 119
 DB 286 EOKRREKEAEER-----QAEERKOKLELETEREKREKREKRLKROKORDRELRRNQ 338
 QY 120 KKV-GISOENKKEVEKTLVPIETKPKFSQ-----AKLT--AGAVKHSSESGNS 167
 DB 339 KKLKLEQAEQKQLOEKIKLE--ERKLLAQNRQISILIAELLISRAKAVKILREGEQKEE 396
 QY 168 VKRLKPDE 176
 DB 397 KLRLOQOE 405

RESULT 14

IF2_AQUAE STANDARD; PRT; 805 AA.

AC 067825;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TRANSLATION INITIATION FACTOR IF-2.
 GN INFB OR AQ_2032.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 RN [1]
 RX SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RA MEDLINE-98196666; PubMed-9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M.A., Keller R.V., Aubay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus."
 RL Nature 392:353-358(1998).
 CC -1- FUNCTION: IF-2, ONE OF THE ESSENTIAL COMPONENTS FOR THE INITIATION
 CC OF PROTEIN SYNTHESIS IN VITRO, PROMOTES FORMYL METHIONYL-TRNA FROM
 CC SPONTANEOUS HYDROLYSIS AND PROMOTES ITS BINDING TO THE 30S
 CC RIBOSOMAL SUBUNITS. IT IS ALSO INVOLVED IN THE HYDROLYSIS OF GTP
 CC DURING THE FORMATION OF THE 70S RIBOSOMAL COMPLEX (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
 CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE000769; AAC0794.1; -;
 DR INTERPRO; IPR000178; -;
 DR INTERPRO; IPR000795; -;
 DR PFM: PFO0009; GTP_EFTU; 1.
 DR PFM: PFO2131; IF2_1.
 DR PROSITE; PS01176; IF2_1.
 KM Initiation factor; Protein biosynthesis; GTP-binding.
 FT DOMAIN 309 458 G-DOMAIN.
 FT NP_BIND 315 322 GTP (BY SIMILARITY).
 FT NP_BIND 362 366 GTP (BY SIMILARITY).
 FT NP_BIND 419 419 GTP (BY SIMILARITY).
 SO SEQUENCE 805 AA; 91600 MW; 8A1BF300C69CA51E CRC64;

Query Match 9.8%; Score 128; DB 1; Length 805;
 Best Local Similarity 23.9%; Pred. No. 1.5;
 Matches 42; Conservative 44; Mismatches 58; Indels 32; Gaps 7;
 QY 10 IKRFRVSAELDERKROE---WEKVRKPEDECEPEVYDPRSLYERLQOKDKO 66

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 17:05:07 ; Search time 13.72 Seconds
(without alignments)
332.441 Million cell updates/sec

Title: US-09-602-597-2
Perfect score: 1306
Sequence: 1 MDGGDGNLIKRRFVEAE.....NATGKIVSIPRTWFLFAP 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/Dackfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 133 | 10.2 | 1162 | 2 | US-08-728-323A-2 |
| 2 | 131.5 | 10.1 | 1898 | 1 | US-08-056-200-94 |
| 3 | 131.5 | 10.1 | 1898 | 2 | US-08-800-644-94 |
| 4 | 125 | 9.6 | 897 | 1 | US-08-095-737-4 |
| 5 | 125 | 9.6 | 897 | 1 | US-08-480-145-4 |
| 6 | 125 | 9.6 | 897 | 2 | US-08-477-388-4 |
| 7 | 123.5 | 9.5 | 1507 | 3 | US-08-929-328-5 |
| 8 | 120.5 | 9.2 | 542 | 1 | US-08-701-380-2 |
| 9 | 120.5 | 9.2 | 542 | 3 | US-09-032-365A-13 |
| 10 | 120 | 9.2 | 432 | 3 | US-08-933-750C-47 |
| 11 | 120 | 9.2 | 432 | 3 | US-09-234-613-47 |
| 12 | 119 | 9.1 | 285 | 1 | US-08-149-809-24 |
| 13 | 118 | 9.0 | 1588 | 4 | PCT-US93-07261-11 |
| 14 | 118 | 9.0 | 1663 | 4 | PCT-US93-07261-16 |
| 15 | 116 | 8.9 | 411 | 2 | US-08-741-134-6 |
| 16 | 115 | 8.8 | 532 | 1 | US-08-285-440-5 |
| 17 | 115 | 8.8 | 532 | 1 | US-08-630-348-5 |
| 18 | 114.5 | 8.8 | 905 | 2 | US-08-574-958A-9 |
| 19 | 114.5 | 8.8 | 1135 | 2 | US-08-574-958A-7 |
| 20 | 113.5 | 8.7 | 717 | 3 | US-08-910-925-1 |
| 21 | 111.5 | 8.5 | 700 | 1 | US-07-720-589-2 |
| 22 | 111.5 | 8.5 | 700 | 2 | US-08-785-190-2 |
| 23 | 111.5 | 8.5 | 700 | 4 | PCT-US92-05539-2 |
| 24 | 110 | 8.4 | 290 | 2 | US-08-903-801-1 |
| 25 | 109.5 | 8.4 | 703 | 3 | US-08-910-925-4 |
| 26 | 109.5 | 8.4 | 2182 | 2 | US-08-487-826B-16 |
| 27 | 109.5 | 8.4 | 3135 | 1 | US-08-323-170B-2 |
| 28 | 108 | 8.3 | 424 | 2 | US-08-951-148-9 |

| | | | | | | |
|----|-------|-----|-----|---|-------------------|--------------------|
| 29 | 108 | 8.3 | 424 | 2 | US-09-165-234-9 | Sequence 9, Appli |
| 30 | 108 | 8.2 | 424 | 3 | US-09-274-570-9 | Sequence 9, Appli |
| 31 | 107.5 | 8.2 | 558 | 1 | US-08-285-440-6 | Sequence 6, Appli |
| 32 | 107.5 | 8.2 | 558 | 1 | US-08-630-348-6 | Sequence 6, Appli |
| 33 | 106.5 | 8.2 | 224 | 2 | US-08-272-255-16 | Sequence 16, Appli |
| 34 | 106.5 | 8.2 | 224 | 4 | PCT-US95-08565-16 | Sequence 16, Appli |
| 35 | 106 | 8.1 | 743 | 2 | US-08-795-475-6 | Sequence 6, Appli |
| 36 | 106 | 8.1 | 743 | 3 | US-08-990-114-3 | Sequence 3, Appli |
| 37 | 105.5 | 8.1 | 714 | 2 | US-09-149-934-4 | Sequence 3, Appli |
| 38 | 105.5 | 8.1 | 746 | 3 | US-08-242-932-2 | Sequence 4, Appli |
| 39 | 105.5 | 8.1 | 984 | 1 | US-08-714-481-2 | Sequence 2, Appli |
| 40 | 105.5 | 8.1 | 984 | 1 | US-08-714-481-2 | Sequence 2, Appli |
| 41 | 105.5 | 8.1 | 984 | 4 | PCT-US95-06111-2 | Sequence 2, Appli |
| 42 | 105 | 8.0 | 288 | 3 | US-08-312-949-4 | Sequence 4, Appli |
| 43 | 105 | 8.0 | 288 | 3 | US-08-446-201-4 | Sequence 4, Appli |
| 44 | 105 | 8.0 | 619 | 1 | US-08-465-746-2 | Sequence 2, Appli |
| 45 | 105 | 8.0 | 619 | 1 | US-08-214-164-2 | Sequence 2, Appli |

ALIGNMENTS

RESULT 1
US-08-728-323A-2
; Sequence 2, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Egelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-728-323A-2

Query Match 10.2% Score 133; DB 2; Length 1162;
Best Local Similarity 25.5% Pred. No. 0.0014;
Matches 41; Conservative 41; Mismatches 61; Indels 18; Gaps 4;
QY 18 EAELDERRRRROEWEKVKRPDPCECEYVDPSLVERLOEOKDRKQGEYEDQFKFN 77

[illegible]

```

1      RESULT      2
2      US-08-056-200-94
3      Sequence 94, Application US/08056200
4      Patent No. 5616500
5      GENERAL INFORMATION:
6      APPLICANT: Steiner, Peter M.
7      APPLICANT: Lee, Seung-Chul
8      APPLICANT: Kim, In-Gyu
9      APPLICANT: Chung, Soo-Il
10     APPLICANT: Park, Sang-Chul
11     TITLE OF INVENTION: Tyriophyllin and Transglutaminase-3 and
12     TITLE OF INVENTION: Methods of Using Same
13     NUMBER OF SEQUENCES: 117
14     CORRESPONDENCE ADDRESS:
15     ADDRESSEE: Knobbe, Martens, Olson & Bear
16     STREET: 620 Newport Center Drive, Sixteenth Floor
17     CITY: Newport Beach
18     STATE: CA
19     COUNTRY: U.S.A.
20     ZIP: 92660
21     COMPUTER READABLE FORM:
22     MEDIUM TYPE: Floppy disk
23     COMPUTER: IBM PC compatible
24     OPERATING SYSTEM: PC-DOS/MS-DOS
25     SOFTWARE: PatentIn Release #1.0, Version #1.25
26     CURRENT APPLICATION DATA:
27     APPLICATION NUMBER: US/08/056,200
28     FILING DATE: 30-APR-1993
29     CLASSIFICATION: 435
30     ATTORNEY/AGENT INFORMATION:
31     NAME: Fedrick, Michael F.
32     REGISTRATION NUMBER: 36,799
33     REFERENCE/DOCKET NUMBER: NIH054.001A
34     TELECOMMUNICATION INFORMATION:
35     TELEPHONE: (714) 760-0404
36     TELEFAX: (714) 760-9502
37     INFORMATION FOR SEQ ID NO: 94:
38     SEQUENCE CHARACTERISTICS:
39     LENGTH: 1898 amino acids
40     TYPE: amino acid
41     TOPOLOGY: linear
42     MOLECULE TYPE: protein
43     US-08-056-200-94

```

```

Query Match          10.1%; Score 131.5; DB 1; Length 1898;
Best Local Similarity 25.8%; Pred. No. 0.0039;
Matches 41; Conservative 35; Mismatches 46; Indels 37; Gaps 5;

QY 12 KRFVSEAEI-----DERRKRQEWKVRKPED-----PECPGEVYDPSLYER 57
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1004 EKRYREBELQOEEOQLLREBEREKRRROGEWERYKKDELQOEEOQLLREBEREKRRLO 1063
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 58 -----LQEQKD-----RKQOEYEQEFKKMNVKGLDETNFLDVSROQL 99
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1064 EROYREBELQOEEOQLLGEERETRROOLEROYRKBELOQOEEOQLLREBEREKRROR 1123
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 100 IEKORREELKE-----LKEYRNNNKKVGIQSENKKEVE 133
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1124 ERQCREEBELQOEEOQLLREBEREKRRROLEQYREEBE 1162
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

RESULT 3
 US-08-800-644-94
 Sequence 94, Application US/08800644
 Patent No. 5358752
 GENERAL INFORMATION:
 APPLICANT: Steinerl, Peter M.
 APPLICANT: Lee, Seung-Chul
 APPLICANT: Kim, In-Gyu
 APPLICANT: Chung, Soo-Il
 TITLE OF INVENTION: Tyrosinase and Tyrosinase-3 and
 NUMBER OF SEQUENCES: 117
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe, Martens, Olson & Bear
 STREET: 620 Newport Center drive, Sixteenth floor
 CITY: Newport Beach
 STATE: CA
 COUNTRY: U.S.A.
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/800,644
 FILING DATE: 14-FEB-1997
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/056,200
 FILING DATE: 30-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Fedrick, Michael F.
 REGISTRATION NUMBER: 36,799
 REFERENCE/DOCKET NUMBER: NIH054.001A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (714) 760-0404
 TELEFAX: (714) 760-9502
 INFORMATION FOR SEQ ID NO: 94:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1898 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-800-644-94

```

Query Match          10.1%; Score 131.5; DB 2; Length 1898;
Best Local Similarity 25.8%; Pred. No. 0.0039;
Matches 41; Conservative 35; Mismatches 46; Indels 37; Gaps 5;

QY      12 KRFVSEAEI-----DEKRRKQEWKVRKPED----PECPCEVYDPSLYER 57
      :::: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1004 EKRYREEEELQOEEOQLLREERERKRRQOEWEQRYKKDELQOEEOQLLREERERKRRQOE 1063

QY      58 -----LQEQKD-----RKQOEYEQKFKMMVRKGLDDETFNFLDEVSRQOEL 99
      :::: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1064 ERQYREEEELQOEEOQLLGEERETRRRQOELERQYKKEBELQOEEOQLLREPEKRRQOE 1123

QY      100 IEKORREEELKE----LKEYRNINLKKGYSOENKKEYE 133
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1124 ERQCREEELQOEEOQLLREERERKRRQOELERQYKKEBEL 1162

RESULT      4
: Sequence 4, Application US/08095737
: Patent No. 5487979
: GENERAL INFORMATION:
: APPLICANT: Difiore, Pier P
: APPLICANT: Fazioli, Francesca

```

TITLE OF INVENTION: A Substrate for the Epidermal Growth
FACTORS
TITLE OF INVENTION: Factor Receptor Kinase
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: California
COUNTRY: United States of America
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,737
FILING DATE: 19930722
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 897 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-095-737-4

Query Match 9.6%; Score 125; DB 1; Length 897;
Best Local Similarity 24.7%; Pred. No. 0.0052;
Matches 54; Conservative 38; Mismatches 71; Indels 56; Gaps 9;
QY 12 KRFRVSE--AELDERKKRQEMEKYRKPEDECEPEYVDRSLYERLOEKDKRQKEY 69
DB 390 QKQVOELLGELDEOKAOLEQLOEVK-----KCAEEAQLISLKAETTSOESQISYE 444
QY 70 EBOFKKNNVGRGDEDETFNLDVSR--QOELJEKORREBELK-ELKEYNNNKKVIGISOE 127
DB 445 EELIKAR-----BELSLQOETAQLESVESGKAQLEPQOHLQOE---SQO 487
QY 128 NKKEVEKLTVPRIETKNKFS-----QAKLLAGAVKH-----KSSSEGN----- 166
DB 488 EISSMGMRLKMDLETNNQNMSSSPQSVLVNGATDYCSLTSSTSETANFNHAGQNN 547
QY 167 -----SVKRLKPDPPDDKNDPSSCKSLGNTSL 196
DB 548 LESEPTHOESSVRSPEIAPSDVTDESEAVTVAGNEKYT 586

RESULT 5
US-08-480-145-4
Sequence 4, Application US/08480145
Patent No. 5717067
GENERAL INFORMATION:
APPLICANT: DiFiore, Pier P
TITLE OF INVENTION: A Substrate for the Epidermal Growth
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: California
COUNTRY: United States of America
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,145
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/095,737
FILING DATE: 22-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 897 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-145-4

Query Match 9.6%; Score 125; DB 1; Length 897;
Best Local Similarity 24.7%; Pred. No. 0.0052;
Matches 54; Conservative 38; Mismatches 71; Indels 56; Gaps 9;
QY 12 KRFRVSE--AELDERKKRQEMEKYRKPEDECEPEYVDRSLYERLOEKDKRQKEY 69
DB 390 QKQVOELLGELDEOKAOLEQLOEVK-----KCAEEAQLISLKAETTSOESQISYE 444
QY 70 EBOFKKNNVGRGDEDETFNLDVSR--QOELJEKORREBELK-ELKEYNNNKKVIGISOE 127
DB 445 EELIKAR-----BELSLQOETAQLESVESGKAQLEPQOHLQOE---SQO 487
QY 128 NKKEVEKLTVPRIETKNKFS-----QAKLLAGAVKH-----KSSSEGN----- 166
DB 488 EISSMGMRLKMDLETNNQNMSSSPQSVLVNGATDYCSLTSSTSETANFNHAGQNN 547
QY 167 -----SVKRLKPDPPDDKNDPSSCKSLGNTSL 196
DB 548 LESEPTHOESSVRSPEIAPSDVTDESEAVTVAGNEKYT 586

RESULT 6
US-08-477-389-4
Sequence 4, Application US/08477389
Patent No. 5872219
GENERAL INFORMATION:
APPLICANT: DiFiore, Pier P
TITLE OF INVENTION: A Substrate for the Epidermal Growth
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: California
COUNTRY: United States of America
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,389

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 542 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-701-380-2

Query Match 9.2%; Score 120.5; DB 1; Length 542;
Best Local Similarity 22.4%; Pred. No. 0.0066;
Matches 55; Conservative 33; Mismatches 107; Indels 51; Gaps 9;

QY 25 RKRROEWE-----KVRKP-----EDPECEPEYVDPKSLYERLQEQDKRQOE 68
DB 42 RKRRTAEPSGCPGSKRRKRGAGRRGRPREPSDPDQAARAPQTVARFLRDEAKRQD 101
QY 69 YEEQFKKNNVRCGLDEDTNPLDEVSROQELIEKOR-----REELKELEKEYRNN 118
DB 102 PREFTVARAPDADEDEEEDDEEEDDEEAEKKEKILLPPKKPLREKSSADLKERRAK 161
QY 119 LKKVIGIOENKKEY-EKKLVKPIETKNKFSQAKLAGAVKHSSESGNSVKKRLKPDPEP 177
DB 162 -----AQGRGDLGSDPPPKPLRVNKK--EAPAGEGTMKRTKKKSGSEADKDPGSGP 213
QY 178 DDKNOEPSSCKSLGNTSLGSPSHCPSAVCGILPGLGAYSGSSDSESSDSEGTI--- 234
DB 214 ASARKSPAMFLVGEXS-----PDKKA---LKKKGTPKGARKKEEEDDEEAAATYIKN 261
QY 235 -NATGK 239
DB 262 SNOKGK 267

RESULT 9

US-09-032-365A-13
Sequence 13, Application US/09032365A
Patent No. 6114502
GENERAL INFORMATION:
APPLICANT: No. 6114502th, Michael
APPLICANT: Nishina, Patsy
APPLICANT: Nagart, Juergen
APPLICANT: No. 6114502en-Trauth, Konrad
TITLE OF INVENTION: GENE FAMILY ASSOCIATED WITH
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Avenue, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,365A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-2C1P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650 327-3231

TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 542 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-032-365A-13

Query Match 9.2%; Score 120.5; DB 3; Length 542;
Best Local Similarity 22.4%; Pred. No. 0.0066;
Matches 55; Conservative 33; Mismatches 107; Indels 51; Gaps 9;

QY 25 RKRROEWE-----KVRKP-----EDPECEPEYVDPKSLYERLQEQDKRQOE 68
DB 42 RKRRTAEPSGCPGSKRRKRGAGRRGRPREPSDPDQAARAPQTVARFLRDEAKRQD 101
QY 69 YEEQFKKNNVRCGLDEDTNPLDEVSROQELIEKOR-----REELKELEKEYRNN 118
DB 102 PREFTVARAPDADEDEEEDDEEEDDEEAEKKEKILLPPKKPLREKSSADLKERRAK 161
QY 119 LKKVIGIOENKKEY-EKKLVKPIETKNKFSQAKLAGAVKHSSESGNSVKKRLKPDPEP 177
DB 162 -----AQGRGDLGSDPPPKPLRVNKK--EAPAGEGTMKRTKKKSGSEADKDPGSGP 213
QY 178 DDKNOEPSSCKSLGNTSLGSPSHCPSAVCGILPGLGAYSGSSDSESSDSEGTI--- 234
DB 214 ASARKSPAMFLVGEXS-----PDKKA---LKKKGTPKGARKKEEEDDEEAAATYIKN 261
QY 235 -NATGK 239
DB 262 SNOKGK 267

RESULT 10

US-08-933-750C-47
Sequence 47, Application US/08933750C
Patent No. 5932442
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purni
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749

RESULT 11
 US-09-234-613-47
 Sequence 47, Application US/09234613
 Patent No. 6132973
 GENERAL INFORMATION:
 APPLICANT: Lal, Preeti
 APPLICANT: Hillman, Jennifer L.
 APPLICANT: Bandman, Olga
 APPLICANT: Shah, Purvi
 APPLICANT: Au-Young, Janice
 APPLICANT: Yue, Henry
 APPLICANT: Guegler, Karl J.
 APPLICANT: Corley, Neil C.
 TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
 NUMBER OF SEQUENCES: 98
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/234,613
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/933,750
 FILING DATE: September 23, 1997
 ATTORNEY/AGENT INFORMATION:

RESULT 12
US-08-149-809-24
Sequence 24, Application US/08149809
Patent No. 5445669
GENERAL INFORMATION:
APPLICANT: METCALFE, Dean D.
APPLICANT: MARTIN, Brian M.
APPLICANT: RAO, Pillarisetti V.S.
TITLE OF INVENTION: IGE-BINDING EPITOPES OF A MAJOR
TITLE OF INVENTION: HEAT-STABLE CRUSTACEAN ALLERGEN DERIVED FROM SHRIMP
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,809
FILING DATE: 10-NOV-1993
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40399/183/NIHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399

TELEX: 904136
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 285 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-149-809-24

Query Match 9.1%; Score 119; DB 1; Length 285;
Best Local Similarity 24.5%; Pred. No. 0.0037;
Matches 53; Conservative 41; Mismatches 80; Indels 42; Gaps 9;

QY 7 GNLIKKRFSEAE-----IDERRKRROEWEK-----YRKRPEDPECPPEYDPRS 53
DB 69 GKLEKKKALONSEVAALNRRLOLEEDLESEERLGSATKLSASQADESERARK 128
QY 54 LYERLOPKDRKQOEYEEQFKNMVGRGLDEDTNPLDEVSRQOELIEK--QRREE---- 107
DB 129 ILERBALADEERBDALENQK---EARFLAEADKKYDEVARKLAWAEADLERAEBAEQ 185
QY 108 -----EL-KEKEYRNKLVKGVISQF--NKKEYEKKLVKPIETKKNKFSQAKLLAGAVK 158
DB 186 GENKIVLELEBELRYVGNLKLSEVSEKSNQREVEYKNOIKTLNTRLKEAEA----- 237
QY 159 HKSESGNSVYKRLKDPD--PDDKNOEPSSCKSLGN 192
DB 238 -RAEFARSVOKLOKEVDRLDLDLVEKERYKDTGD 272

RESULT 13
PCT-US93-07261-11

; Sequence 11, Application PC/TUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION: PFEPM3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1588 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-07261-11

Query Match 9.0%; Score 118; DB 4; Length 1588;
Best Local Similarity 24.7%; Pred. No. 0.048;
Matches 58; Conservative 40; Mismatches 67; Indels 70; Gaps 13;

QY 5 DDGNLIKK-----RFVSEALDER-----RKRROEWM-----EKVRKPED 40
DB 105 DDKKIYFKKPESEYKDLKEKELKEKFIKOHLDYERKKRRNNILRSIRROKLEIEQ 164
QY 41 PE-----ECPEEYDPRSLYERLO-----EOKRKQOEYEEQFKKNVGRGLDED 85
DB 165 LKTLNQLSALINELKERRASRRPMVYKMGKGVDEVDWIKKYDDEQAKN---GTKDE 221
QY 86 ETNPLDEVSRQOELIEKQ---RREBELKEKEY-----RNNLKVGISQENKKEYEKK 135
DB 222 EIK--DKGDYEIEVEIKFYGMENALGELDEVEEYERKKRYLKEGD--EGDLKDYEEK 277
QY 136 LTVKPIETKKNKFSQAKLLAGAVKHKSESGNSVYKRLKPD-----PEPDDKNQ 182
DB 278 LEETGGERKEKFTTRL--YRKRRK--EQKKLKEDEKKLLAAEPPDERK 326

RESULT 14
PCT-US93-07261-16

; Sequence 16, Application PC/TUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION: PFEPM3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORGANISM: Plasmodium falciparum
; STRAIN: Malayan Camp
PCT-US93-07261-16

Query Match 9.0%; Score 118; DB 4; Length 1663;
Best Local Similarity 24.7%; Pred. No. 0.052;
Matches 58; Conservative 40; Mismatches 67; Indels 70; Gaps 13;

QY 5 DDGNLIKK-----RFVSEALDER-----RKRROEWM-----EKVRKPED 40
DB 105 DDKKIYFKKPESEYKDLKEKELKEKFIKOHLDYERKKRRNNILRSIRROKLEIEQ 164
QY 41 PE-----ECPEEYDPRSLYERLO-----EOKRKQOEYEEQFKKNVGRGLDED 85
DB 165 LKTLNQLSALINELKERRASRRPMVYKMGKGVDEVDWIKKYDDEQAKN---GTKDE 221

OY 86 ETNFDENVSRQOELIEKQ---RREBELKELKEY-----RNNLKKVGISQENKKEVEKK 135
DB 222 EIK--DKGGGEIYEIVETLKYCGMRENALGELDEYEEYERKRYLLKEDG--EGDLKDVEEK 277
OY 136 LTVKPIETKFKRSQAKILAGAVKHKSSSGNSVRLKPD-----PEPDCKNO 182
DB 278 LEETGYGFRKFPPTTRIL--VKRRNK---EOKKLKEDKELKLIABEPDDEKK 326

RESULT 15

US-08-741-134-6
; Sequence 6, Application US/08741134
; Patent No. 5861498
; GENERAL INFORMATION:
; APPLICANT: Litwack, Gerald
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandes-Alnemri, Teresa
; TITLE OF INVENTION: IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
; TITLE OF INVENTION: AND
; TITLE OF INVENTION: METHODS OF USING THE SAME
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & NO. 5861498rls
; STREET: One Liberty Place - 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 3.11
; SOFTWARE: Wordperfect for Windows 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/741,134
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,163
; FILING DATE: 01-NOV-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-2090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-741-134-6

Query Match 8.9%; Score 116; DB 2; Length 411;

Best Local Similarity 23.0%; Pred. No. 0.011; Mismatches 79; Indels 76; Gaps 9;

OY 3 GDDDNLLIKKRFVSEALDERRRKROEWEKVRK-----PEDPECEPE--- 47
DB 68 GSD-----PDEDEIDESSSEEEB-EKTOKKKKSKGKKAESSEDEDEDEDE 116
OY 48 -----VYDPRSLYERLOEOAKRKOQYE-----EOKFKNNVREGID 83
DB 117 FOESVLLTLSPAOYQOSIDLITTPREEVQFVTSYAISLSGNYVKHFDTPMGVEGED 176
OY 84 EDE-----TNFDENVSRQOELI-----EKORREBELKELKEYRNNLKKVGISQ 126
DB 177 EDEADADYDSEYDILTPDEDEITIGDDMDLDDDEEEVEVRIEVOEDEDENDGE---EQ 233

OY 127 ENKKEVEKKLTVKPIETKFKRSQAKILAGAVKHKSSSGNSVRLKPDPEPDCKNOEPPSS 186
DB 234 EEEEEEEQKEEVKPPPKSKKKEKRR-----KHEEKEEERAKAKVKVYFKKDLBEGPTK 287
OY 187 CKS 189
DB 288 PKs 290

Search completed: February 16, 2001, 17:06:52
Job time: 105 sec

~~BEST AVAILABLE COPY~~

THIS PAGE BLANK (UST)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 17:05:07 ; Search time 25.95 Seconds
(without alignments)
1147.237 Million cell updates/sec

Title: US-09-602-597-2
Perfect score: 1306
Sequence: 1 MDGGDDGNLTKRRFVSEAF.....NATGKIVSSIRPTWFLFAP 254

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_15:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp_invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp rodent:*
12: sp.virus:*
13: sp vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 422.5 | 32.4 | 249 | 5 | 09V843 |
| 2 | 302.5 | 23.2 | 219 | 5 | 017594 |
| 3 | 169.5 | 13.0 | 241 | 10 | 09M1R0 |
| 4 | 160.5 | 12.3 | 2274 | 5 | 09VYU0 |
| 5 | 151.5 | 11.6 | 3476 | 5 | 09NM17 |
| 6 | 151.5 | 11.6 | 5533 | 5 | 09U6C3 |
| 7 | 151.5 | 11.6 | 5554 | 5 | 09NMN1 |
| 8 | 151 | 11.6 | 651 | 4 | 09UPZ5 |
| 9 | 150.5 | 11.5 | 693 | 5 | 09VNG4 |
| 10 | 149.5 | 11.4 | 312 | 5 | 09V570 |
| 11 | 149 | 11.4 | 1087 | 5 | 09G923 |
| 12 | 148 | 11.3 | 314 | 5 | 015987 |
| 13 | 147 | 11.3 | 706 | 11 | 09QX56 |
| 14 | 146 | 11.2 | 271 | 5 | 044077 |
| 15 | 146 | 11.2 | 293 | 5 | 015988 |
| 16 | 146 | 11.2 | 451 | 2 | 050870 |
| 17 | 146 | 11.2 | 460 | 2 | 09ZIU2 |
| 18 | 145.5 | 11.1 | 699 | 10 | 09ZU69 |
| 19 | 145.5 | 11.1 | 2701 | 4 | 09Y520 |

| | | | | | | |
|----|-------|------|------|----|--------|--------------------|
| 20 | 144.5 | 11.1 | 513 | 10 | 09LW95 | 091W95 nleotiana t |
| 21 | 143 | 10.9 | 661 | 11 | 070205 | 070205 rattus norv |
| 22 | 143 | 10.9 | 1165 | 4 | 095819 | 095819 homo sapien |
| 23 | 143 | 10.9 | 2081 | 10 | 09LH98 | 091H98 arabidopsis |
| 24 | 142 | 10.9 | 772 | 13 | 091013 | 091013 gallus gall |
| 25 | 142 | 10.9 | 1661 | 5 | 006166 | 006166 plasmodium |
| 26 | 141.5 | 10.8 | 384 | 5 | 09XZ71 | 09XZ71 periplaneta |
| 27 | 141.5 | 10.8 | 466 | 5 | 09J312 | 09J312 caenorhabd |
| 28 | 141 | 10.8 | 1027 | 4 | 09NS77 | 09NS77 homo sapien |
| 29 | 141 | 10.8 | 1257 | 4 | 09S033 | 09S033 homo sapien |
| 30 | 140.5 | 10.8 | 1175 | 4 | 075172 | 075172 homo sapien |
| 31 | 140 | 10.7 | 380 | 5 | 09XZ72 | 09XZ72 libellula p |
| 32 | 139.5 | 10.7 | 1183 | 11 | 089040 | 089040 rattus norv |
| 33 | 139 | 10.6 | 1233 | 11 | P97820 | P97820 mus musculu |
| 34 | 138.5 | 10.6 | 335 | 5 | 076719 | 076719 caenorhabd |
| 35 | 138 | 10.6 | 385 | 5 | 017909 | 017909 caenorhabd |
| 36 | 137 | 10.5 | 1359 | 5 | 002061 | 002061 caenorhabd |
| 37 | 137 | 10.5 | 1359 | 5 | 09U7E0 | 09U7E0 caenorhabd |
| 38 | 136.5 | 10.5 | 673 | 5 | 09U0N1 | 09U0N1 plasmodium |
| 39 | 136 | 10.4 | 660 | 11 | 09QX55 | 09QX55 mus musculu |
| 40 | 135 | 10.3 | 729 | 3 | 09P3P1 | 09P3P1 neosporea |
| 41 | 135 | 10.3 | 777 | 3 | 045373 | 045373 caenorhabd |
| 42 | 135 | 10.3 | 919 | 4 | 09NQ57 | 09NQ57 homo sapien |
| 43 | 134.5 | 10.3 | 1010 | 5 | 09VB01 | 09VB01 drosophila |
| 44 | 134 | 10.3 | 554 | 5 | 09VPS3 | 09VPS3 drosophila |
| 45 | 133.5 | 10.2 | 735 | 5 | 09U0D5 | 09U0D5 tetrahymena |

ALIGNMENTS

RESULT 1
ID 09V843 PRELIMINARY; PRT; 249 AA.
AC 09V843:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE CG14480 PROTEIN.
GN CG14480.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brockett P., Brothier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster A., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegyam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon R., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shne B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003802; AAF57835.1; -
 DR FLYBASE: FBgn0034242; CG14480.
 SQ SEQUENCE 249 AA: 28376 MW: 070FA9AE2764849E CRC64:

Query Match 32.4%; Score 422.5; DB 5; Length 249;
 Best Local Similarity 41.8%; Pred. No. 4.2e-18;
 Matches 94; Conservative 44; Mismatches 76; Indels 11; Gaps 5;

QY 15 FVSEELDERRRKROEEMKVKRPDPPECEPEYVDPSLVERLOEQDKROQEEBQK 74
 DB 5 FVTEAEAEQROREWEVERQRPEDPLEPEPYDGRSLVERLKNQDKMEFEFAHK 64
 QY 75 FKNMWVGDEDETNFLDEYSRQOELIEKORREELKELKEYNNKVKYGISOE--EVE 133
 DB 65 LKNLRIGLDDDEVQPLEYDAHKIHAERQOMDELELKFNNRREYKLOESVDDKLOE 124
 QY 134 KKLIVKPI-ETKKNFSQAKLLAGAVKHKSSSGNSVRLK-----PPPEDDKNQEPSS 186
 DB 125 LKTTAKSAGASVGRSTQGLLGQIKRKNKNGELPTTSKVAKITENEVEQTATNEATKPAD 184
 QY 187 CKSLGNTSLSGPSIHCSAAVCIIGLPGAGVSSGSDSESSDSE 231
 DB 185 -KTTWTLITNKK--YNOGALKCIAILPGISYSSSDSEASTDEE 226

RESULT 2
 017594 PRELIMINARY; PRT: 219 AA.
 ID 017594;
 AC 017594;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
 DE C25A1.1 PROTEIN.
 GN C25A1.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitidae;
 OC Rhabdilitidae; Peloderiinae; Caenorhabditis.
 OC NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mortimore B.;
 RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-94150718; PubMed-7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Boxfield M., Dear S., Du Z., Durbin R., Faveall A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*
 RT elegans.";
 RL Nature 368:32-38(1994).

DR EMBL: 281038; CAB02762.1; -
 SQ SEQUENCE 219 AA: 24353 MW: 9DD37690588FDEB5 CRC64:

Query Match 23.2%; Score 302.5; DB 5; Length 219;
 Best Local Similarity 36.4%; Pred. No. 4.4e-11;
 Matches 82; Conservative 36; Mismatches 70; Indels 37; Gaps 8;

QY 15 FVSEELDERRRKROEEMKVKRPDPPECEPEYVDPSLVERLOEQDKROQEEBQK 74
 DB 5 FVTEAEAEQROREWEVERQRPEDPLEPEPYDGRSLVERLKNQDKMEFEFAHK 64
 QY 75 FKNMWVGDEDETNFLDEYSRQOELIEKORREELKELKEYNNKVKYGISOE--NKEEV 132
 DB 65 LKNMWVGDEDESVFLSELSTKRVV-KMRKKEEQEL-----IKELAVQHIAANPS 117
 QY 133 EKKLIVKPIETK---NFSQAKLLAGAVKHKSSSGNSVRLKADPEDDKNQEPSSCK 188
 DB 118 SSRFLIKPSTSKVLGPSPKQAFLSTAIKRKSTST-----EERKOE-- 159
 QY 169 SLGNTSLSGPSIHCSAAVCIIGLPGAGVSSGSDSE--SSDSE 231
 DB 160 DVSSKSKPE---PVTKQIGALQALCEIDPSSDSESDASSDDE 200

RESULT 3
 09M1R0 PRELIMINARY; PRT: 241 AA.
 ID 09M1R0;
 AC 09M1R0;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE HYPOTHETICAL 27.2 KDA PROTEIN.
 GN T17J13.100.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Mewes H.W.,
 RA Lemcke K., Mayer K.F.X., Quetler F., Salanoubat M.;
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL138651; CAB71874.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 241 AA: 27168 MW: 350F997D29307F14 CRC64:

Query Match 13.0%; Score 169.5; DB 10; Length 241;
 Best Local Similarity 23.3%; Pred. No. 0.0034;
 Matches 63; Conservative 30; Mismatches 78; Indels 99; Gaps 8;

QY 15 FVSEELDERRRKROEEMKVKRPDPPECEPEYVDPSLVERLOEQDKROQEEBQK 74
 DB 14 FVSEQDDESKRERGERVEDGTFRD-----RALTEILKENQDKDAEFNEKFK 62
 QY 75 FKNMW-----RGDEDETNFLDEYSRQOELIEKORREEE 108
 DB 63 HSESLSTVSSIDSYSCDQFLCGLCGPPKALDEDETEFLDKI----- 107
 QY 109 LKEIKYNNKLVKYGISOEKNKEVEKKITVPPIETKNSQAKLLAGAVKHKSSSGNSV 168
 DB 108 -----EMSKREYERQLANEDEDQLRNFQAQVAARSAITLHPKREA--- 147
 QY 169 KRLKPDPEDDKNQEPSSCKSLGNTSLSGPSIHCSAAVCI-----GLPG 214
 DB 148 ---LPPAPVTKQKP-----IGKRN---PATRPFKALIVKPPKAKATEKEKEKEITPG 196

QY 215 LGAYSSSDSSSSDEGTINATKIYSSI 244
DB 197 NKGASHIDQASIDSVKG--NWTGKTTEGL 224

RESULT 4
Q9YU00 PRELIMINARY; PRT; 2274 AA.

AC Q9YU00;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE CG1905 PROTEIN.
GN CG1905.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC NCB1_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mklos G.L.G.,
RA Abril J.F., Adayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Bussan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spires E., Spirdling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtk R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wessertman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003487; AAF48098.1; -
DR FLIBASE: FBgn0030337; CG1905.
SQ SEQUENCE 2274 AA; 24709 MW; B287311CBDF9EF20 CRC64;

Query Match 12.3%; Score 160.5; DB 5; Length 2274;
Best Local Similarity 28.8%; Pred. No. 0.11;
Matches 69; Conservative 34; Mismatches 82; Indels 55; Gaps 10;

QY 11 IKRPVSEALDRRRKROEWEKVRKPEDEPCPE---EYVDPSTLERTLOEQDRKO 66
DB 1477 LREKIKEREKE--KLKREERKMKREKREKIKEREKEREKREKREKREKREK 1534

QY 67 QEYEQFKFKNMVRGLDEDTNLFDEVSROQELIEKOR-----REBELK 110
DB 1535 KEKEELKKKEKE---EKEKEELKEREKROEKEKREKREKREKREKREKREK 1591

QY 111 ELKEVRNNLK-----KVGISQENKKEVEKRLTKVPIETKKNFSQAKLLAGAVKRRSSS 164
DB 1592 E-KERAEKLLDEKEVKLKEKEEQGLKEKEKELKKEKKEKEDYKE-----KESLES 1642

QY 165 -----GNSYKRLKPPDPDDKNOEBSKSLCNTSLTSGPSIHCPAANCIGILPGL 215
DB 1643 EKLLISATVSNPWRVYEDPPP--KLPAVQDYSPLGK---PTKASPEKRDKEKLPGL 1696

RESULT 5
Q9NU17 PRELIMINARY; PRT; 5476 AA.

AC Q9NU17;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE SPLIT ENDS.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC NCB1_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Rebay J., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Laverly T.,
RA Sun C., Voas M., Williams A., Rubin G.M.;
RT "A Genetic Screen for Novel Components of the Ras/Mitogen-Activated
RT Protein Kinase Signaling Pathway That Interact With the Yan Gene of
RT Drosophila Identifies Split Ends, a New RNA Recognition Motif-
RT Containing Protein.";
RL Genetics 154:695-712(2000).
DR EMBL: AF184612; AAF26299.1; -
SQ SEQUENCE 5476 AA; 590531 MW; 93FA8C7860770C2 CRC64;

Query Match 11.6%; Score 151.5; DB 5; Length 5476;
Best Local Similarity 24.2%; Pred. No. 0.89;
Matches 46; Conservative 44; Mismatches 87; Indels 13; Gaps 2;

QY 18 EALDER---RRROEWEKVRKPEDEPCPEYVDPSTLERTLOEQDRKOEEYEQF 73
DB 1893 KAOQEEREDKREKREKREKREKREKREKREKREKREKREKREKREKREKREK 1952

QY 74 KKKNNVRGIDDEDTNLFDEVSROQELIEKORREBELKELKEVRNNLKVGISQENKKEVE 133
DB 1953 LREKEQEREDNREKRLDRDKLREKREKREKREKREKREKREKREKREKREK 2003

QY 134 KRLTVPIETKKNFSQAKLLAGAVKRRSSSGNSYKRLKPPDPDDKNOEBSKSLGNT 193
DB 2004 QSRRAADVQEGRGGMRELSSYQKSKMDIAGASSLTALDOHNKNNAMDITAGSTPPA 2063

QY 194 SLGSPSINCP 203
DB 2064 SPSTPSDNT 2073

RESULT 6
Q9U6C3 PRELIMINARY; PRT; 5533 AA.

AC Q9U6C3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE SPEN RNP MOTIF PROTEIN LONG ISOFORM.
GN SPEN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miles G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Plankkoch C., Baldwin D.,
 RA Balles R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle C., Daveport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fliselman W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Moberly C., Morris J., Moshireli A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murty D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Rehner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs S.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS: LONG ISOFORM (SHOWN HERE) AND
 CC SHORT ISOFORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
 DR EMBL: A0003602; AAF51972.1; -;
 DR EMBL: A0003602; AAF51973.1; -;
 DR FLYBASE: FBgn0037362; CG2179.
 KW Alternative splicing; Hypothetical protein.
 FT VARSPLIC 1 135 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 693 AA; 80212 MW; P65BA9601F086EDD CRC64;

Query Match 11.5%; Score 150.5; DB 5; Length 693;
 Best Local Similarity 23.1%; Pred. No. 0.13; Mismatches 83; Indels 61; Gaps 8;
 Matches 58; Conservative 49;

QY 11 IKRREYSEALDERRRKROEEMKVPEDPECEPEEYVDPSLYERLOQORQOEYE 70
 DB 277 IRRKFFAKA-----KAEDEREKLRKQOE-----ELDERQKOEELKAEAKERERE 325
 QY 71 EOKFPMVNRGLDEDFENFLDEYS--RQOELEIKORREELKEK-----EYNNNLKTVG 123
 DB 326 ERKKEKHKMKIQERGVEISQKIRIEKRLIAQRKLESLITLLEKLFERITQLQORKNRR 385
 QY 124 ISOENKKEVEKLTVPRIETKKNFSQAKL-----LAGAVKHKSSSGNS 167
 DB 386 VKGDDDDKDIORGLVE-----KYKAATEKLVCDOKRIVQELKSWTPPLGLLKSCKSSKSA-- 439
 QY 168 VKRLAKDPEDDKNQEPSSCKSLGNTSLSGPSIHCPS-----AAVCIGILPG 214
 DB 440 -----PVDASSEDEAAASKRRHMGNGVTESAPPGVNSALNPLNVAAMAAGAVRG 491
 QY 215 LGAYSGSSDE 225
 DB 492 PGGYSAKAVASE 502

RESULT 10

P91570
 ID P91570 PRELIMINARY; PRT; 312 AA.
 AC P91570;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE COSMID ZK354.
 GN ZK354.3
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; *Caenorhabditis*.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=790638;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showmken R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.;
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Johnson D., Wamsley P., Bradshaw H.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U081172; AAB42258.1; -;
 SQ SEQUENCE 312 AA; 36737 MW; 958C146E0021E42E CRC64;

Query Match 11.4%; Score 149.5; DB 5; Length 312;
 Best Local Similarity 24.9%; Pred. No. 0.066; Mismatches 78; Indels 19; Gaps 5;
 Matches 46; Conservative 42;

QY 20 ELDERRRKROEEMKVPEDPECEPEEYVDPS-----LYERLOQORQOE 66
 DB 68 EKKEEKEEKEEKEEKKADDEKKTTEEDKDKSKTEEDKTSVKTQTKSERDKKD 127
 QY 67 QEYEEQFKFNWVRGLDEDFENFLDEV--SRQOELEIKORREELKEKRYNNLKRVGI 124
 DB 128 ERKDEKKEEENKESKDEEKK--DEVYKDKDEKMDKKPKGEKEEKEEKKEMKEEK 185
 QY 125 SQENKKEVEK-LTVPRIETKKNFSQAKLAGAVKHKSSSGSVRLKADPPEDDKNOE 183
 DB 186 KEKKEEKKPKNDAPKKEGTGKGVKDALVENDIVTEMDSR-DEKKEKKDKDKDEKDE 244
 QY 184 PSSCK 188
 DB 245 KKEAK 249

RESULT 11
 ID 096923 PRELIMINARY; PRT; 1087 AA.
 AC 096923;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE GELSOLIN-RELATED PROTEIN.
 GN GRP120.

THIS PAGE BLANK